Nonparametric inference for regression models with spatially correlated errors

Andrea Meilán Vila

PhD Thesis

2020

UNIVERSIDADE DA CORUÑA

Nonparametric inference for regression models with spatially correlated errors

Author: Andrea Meilán Vila

Supervisors: Mario Francisco Fernández and Rosa M. Crujeiras

PhD Thesis in Statistics and Operations Research

Department of Mathematics

Universidade da Coruña

The undersigned certify that they are the advisors of the Doctoral Thesis entitled "Nonparametric inference for regression models with spatially correlated errors", developed by Andrea Meilán Vila at the University of A Coruña (Department of Mathematics), as part of the interuniversity PhD program (UDC, USC and UVigo) of Statistics and Operational Research, opting to the International PhD Mention, and hereby give their consent to the author to proceed with the thesis presentation and the subsequent defense.

Os abaixo asinantes fan constar que son os directores da Tese de Doutoramento titulada "Nonparametric inference for regression models with spatially correlated errors", desenvolta por Andrea Meilán Vila na Universidade da Coruña (Departamento de Matemáticas) no marco do programa interuniversitario (UDC, USC e UVigo) de doutoramento en Estatística e Investigación Operativa, optando á Mención Internacional de Doutoramento, e dan o seu consentimento para que a autora proceda á súa presentación e posterior defensa.

Los abajo firmantes hacen constar que son los directores de la Tesis Doctoral titulada "Nonparametric inference for regression models with spatially correlated errors", realizada por Andrea Meilán Vila en la Universidade da Coruña (Departamento de Matemáticas) en el marco del programa interuniversitario (UDC, USC y UVigo) de doctorado en Estadística e Investigación Operativa, optando a la Mención Internacional de Doctorado, y dan su consentimiento para que la autora proceda a su presentación y posterior defensa.

A Coruña, 18 November, 2020

Advisors:

FRANCISCO FERNANDEZ MARIO - 32660323R

Firmado digitalmente por FRANCISCO FERNANDEZ MARIO - 32660323R
Fecha: 2020.11.26 12:59:30 +01'00'

CRUJEIRAS CASAIS ROSA MARIA - 52931780V

Firmado digitalmente por CRUJEIRAS CASAIS ROSA MARIA - 52931780V
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-52931780V, givenName=ROSA MARIA, sn=CRUJEIRAS CASAIS, cn=CRUJEIRAS CASAIS ROSA MARIA - 52931780V
Fecha: 2020.11.26 13:03:36 +01'00'

Dr. Mario Francisco Fernández

Dra. Rosa M. Crujeiras Casais

PhD Student:

MEILAN VILA ANDREA - 33348893C

Firmado digitalmente por MEILAN VILA ANDREA - 33348893C
Nombre de reconocimiento (DN): c=ES, serialNumber=IDCES-33348893C, givenName=ANDREA, sn=MEILAN VILA, cn=MEILAN VILA ANDREA - 33348893C
Fecha: 2020.11.26 12:30:53 +01'00'

Andrea Meilán Vila

A mis padres y a mi hermana

Un viaje de mil millas comienza con un primer paso

Lao-tsé

Agradecimientos

En primer lugar, quiero expresar mi más profundo agradecimiento a mis directores de tesis, Mario Francisco Fernández y Rosa M. Crujeiras, por el apoyo recibido a lo largo de estos años, tanto en lo profesional como en lo personal. A ellos les debo la madurez investigadora que he adquirido. A Mario le agradezco que siempre haya tenido un momento para mí, sus profundas y exhaustivas revisiones y la continua motivación que me ha transmitido. A Rosa por haberme animado a dar mis primeros pasos en la investigación, por guiarme durante este camino y por todas sus críticas constructivas.

I would like to show my gratitude to Jean Opsomer for his hospitality and guidance in research during my stay at Fort Collins. I would like also to thank Agnese Panzera for her warm welcome during my stay in Florence and for her wonderful supervision during that time. Thanks to Anneleen Verhasselt to provide me the opportunity of visiting the Center for Statistics at Hasselt and conducting research.

Gracias a Wenceslao González Manteiga por la confianza depositada en el inicio de mi etapa en la investigación con mi primer contrato laboral. A *Wences* también quiero agradecerle sus comentarios y sugerencias, especialmente en el primer periodo de la tesis. Gracias también a Rubén Fernández Casal por acompañarme en la senda de la programación.

A mis padres, a vosotros os lo debo todo. Gracias por vuestro cariño, por el apoyo incondicional y por ser un ejemplo de fuerza y superación. A mi hermana, por sacarme una sonrisa en mis peores momentos, por tu apoyo constante y por ser mi mejor compañera de vida.

Finalmente, me gustaría darle las gracias a mis amigos, esa familia que escogemos. A Sara, Adolfo, Cris, Miriam, Suso, Xabi, Franco, Paola y Sara, por haber estado

ahí a lo largo de estos años. Gracias a Toni por el *día a día*, me has ayudado a desconectar cuando era necesario. También quiero darle las gracias a Inés, Guido y Silvia, por el apoyo y la compañía diaria. Gracias también a todos los compañeros de doctorado, especialmente a Lucía, Arís y María.

En definitiva, gracias a todos por haberme acompañado durante esta etapa.

A Coruña, 18 de Noviembre de 2020
Andrea Meilán Vila

This research has been partially supported by MINECO Grants MTM2014-52876-R and MTM2017-82724-R, and by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015 and ED431C-2020-14, and Centro de Investigación de Galicia "CITIC" ED431G/01 and ED431G 2019/01), all of them through the ERDF (European Regional Development Fund). Financial support from the Xunta de Galicia Grant ED481A-2017/361 and the European Union (European Social Fund - ESF), is gratefully acknowledged. Part of the results of this thesis have been obtained during a stay of the author at Colorado State University, financed by the Xunta de Galicia Grant ED481A-2017/361, and by the MINECO Grant MTM2014-52876-R, and during the visits to the Università degli Studi di Firenze and to the Universiteit Hasselt, financed by the Xunta de Galicia (Grupos de Referencia Competitiva ED431C-2016-015).

Abstract

Regression estimation can be approached using nonparametric procedures, producing flexible estimators and avoiding misspecification problems. Alternatively, parametric methods may be preferable to nonparametric approaches if the regression function belongs to the assumed parametric family. However, a bad specification of this family can lead to wrong conclusions. Regression function misspecification problems can be somewhat tackled by applying a goodness-of-fit test. For data presenting some kind of complexity, for example, circular data, the approaches used in regression estimation or in goodness-of-fit tests have to be conveniently adapted. Moreover, it might occur that the variables of interest can present a certain type of dependence. For example, they can be spatially correlated, where observations which are close in space tend to be more similar than observations that are far apart. The goal of this thesis is twofold, first, some inference problems for regression models with Euclidean response and covariates, and spatially correlated errors are analyzed. More specifically, a testing procedure for parametric regression models in the presence of spatial correlation is proposed. The second aim is to design and study new approaches to deal with regression function estimation and goodness-of-fit tests for models with a circular response and an $\mathbb{R}^d$-valued covariate. In this setting, nonparametric proposals to estimate the circular regression function are provided and studied, under the assumption of independence and also for spatially correlated errors. Moreover, goodness-of-fit tests for assessing a parametric regression model are presented in these two frameworks. Comprehensive simulation studies and application of the different techniques to real datasets complete this dissertation.

Keywords: circular statistics, goodness-of-fit test, linear-circular regression, nonparametric estimation, spatial dependence

Resumo

A estimación da regresión pode ser abordada empregando técnicas non paramétricas, dando lugar a estimadores flexibles e evitando problemas de mala especificación. Alternativamente, os métodos paramétricos poden ser preferibles se a función de regresión pertence á familia paramétrica asumida. Porén, unha mala especificación desta familia pode levar a conclusións equivocadas. Os problemas de especificación incorrecta da función de regresión poden ser abordados aplicando un contraste de bondade de axuste. Para datos que presentan algún tipo de complexidade, por exemplo, datos circulares, os métodos empregados na estimación ou nos contrastes, deben adaptarse convenientemente. Ademais, pode ocorrer que as variables de interese poidan presentar un certo tipo de dependencia. Por exemplo, poden estar espacialmente correladas, onde as observacións que están preto no espazo tenden a ser máis similares que as observacións que están lonxe. O obxectivo desta tese é dobre, primeiro, analízanse problemas de inferencia para modelos de regresión con resposta e covariables Euclídeas, e erros espacialmente correlados. Máis concretamente, contrástase se a función de regresión pertence a unha familia paramétrica, en presenza de correlación espacial. O segundo obxectivo é deseñar e estudar novos procedementos para abordar estimación e contrastes da función regresión para modelos con resposta circular e covariable con valores en $\mathbb{R}^d$. Neste contexto, preséntanse e estúdanse propostas non paramétricas para estimar a función de regresión circular, baixo o suposto de independencia e tamén para erros espacialmente correlados. Ademais, nestes dous contextos, preséntanse contrastes para avaliar un modelo de regresión paramétrico. Esta memoria complétase con estudos de simulación exhaustivos e aplicacións a conxuntos de datos reais.

Palabras clave: contraste de bondade de axuste, estatística circular, estimación non paramétrica, regresión lineal-circular, dependencia espacial

Resumen

La estimación de la regresión puede ser abordada usando técnicas no paramétricas, dando lugar a estimadores flexibles y evitando problemas de mala especificación. Alternativamente, los métodos paramétricos pueden ser preferibles si la función de regresión pertenece a la familia paramétrica asumida. Sin embargo, una mala especificación de esta familia puede llevar a conclusiones equivocadas. Los problemas de especificación incorrecta de la función de regresión pueden ser abordados aplicando un contraste de bondad de ajuste. Para datos que presentan algún tipo de complejidad, por ejemplo, datos circulares, los métodos utilizados en la estimación o en los contrastes, deben adaptarse convenientemente. Además, puede ocurrir que las variables de interés puedan presentar un cierto tipo de dependencia. Por ejemplo, pueden estar espacialmente correladas, donde las observaciones que están cerca en el espacio tienden a ser más similares que las observaciones que están lejos. El objetivo de esta tesis es doble, primero, se analizan problemas de inferencia para modelos de regresión con respuesta y covariables Euclídeas, y errores espacialmente correlados. Más concretamente, se contrasta si la función de regresión pertenece a una familia paramétrica, en presencia de correlación espacial. El segundo objetivo es diseñar y estudiar nuevos procedimientos para abordar estimación y contrastes de la función regresión para modelos con respuesta circular y covariable con valores en $\mathbb{R}^d$. En este contexto, se presentan y estudian propuestas no paramétricas para estimar la función de regresión, bajo el supuesto de independencia y también para errores espacialmente correlados. Además, en estos dos contextos, se presentan contrastes para evaluar un modelo de regresión paramétrico. Esta memoria se completa con estudios de simulación exhaustivos y aplicaciones a conjuntos de datos reales.

Palabras clave: contraste de bondad de ajuste, estadística circular, estimación no paramétrica, regresión lineal-circular, dependencia espacial

Contents

Chapter 1

Introduction

The analysis of a variable of interest which depends on other variable(s) is a classic problem that appears in many disciplines. To tackle this issue, an appropriate regression model setting up the possible functional relationship between the variables is usually formulated. A key element of such a model is the regression function, describing the general relationship between the variable of interest (response or dependent variable) and the explanatory variable(s) (covariates, predictors or independent variables). Once the regression function (which is usually unknown) is estimated, and the model is properly validated, values of the response variable for known values of the covariate(s) could be predicted, for instance.

In the present dissertation, models with univariate response variable and several covariates, presenting a certain type of spatial dependence structure, will be considered. In this situation, both the regression function and the dependence structure must be appropriately specified. Modeling properly the regression function and searching for suitable regression models is crucial to obtain reliable estimations and predictions.

For regression models with Euclidean response and covariates, and spatially correlated errors, using parametric approaches, regression estimation can be carried out using least squares tools or maximum likelihood methods, as described in Diggle et al. (2010). On the other hand, without assuming a parametric form for the regression function, nonparametric regression estimation methods, for example kernel-type approaches, can also be employed in this context. The asymptotic properties of

different kernel estimators were derived in Liu (2001), in the setting of multiple regression with spatially correlated errors and short-range dependence. Therefore, an important question arising when estimating the regression function is if a parametric or a nonparametric approach should be used. If the form of this function is known (at least partially), then parametric methods may be used to model the dataset properly. However, if a wrong functional form is supposed for this function, then inaccurate estimations will be obtained, and comparing with other estimating methods, with larger bias (Fan and Yao, 2008). Despite parametric models are easy to work with, nonparametric methods relax parametric assumptions, and consequently enable one to explore the data more flexibly. Regression function misspecification problems can be somewhat addressed by applying a goodness-of-fit test. Although there is a substantial literature on assessing a parametric regression model, confronting a parametric estimator of the regression function with a smooth alternative estimated by nonparametric procedures (Alcalá et al., 1999; Azzalini et al., 1989; Eubank and Spiegelman, 1990; González Manteiga and Vilar Fernández, 1995; Härdle and Mammen, 1993), this is not the case for spatially correlated data.

When data present certain complexities, classical regression procedures designed for Euclidean data could not be directly employed. This is the case, for instance, when working with circular data. Circular data are represented as points on the circumference of a unit circle. Observations from circular processes are quite frequent in applied sciences such as, oceanography, meteorology, or biology, among others. Due to the angular nature of such data, some difficulties can be found trying to apply traditional statistical methods. For example, considering a multiple linear-circular regression model in which a random angle may depend on several real-valued random covariates, the circular regression function is given by the inverse tangent function of the ratio between the conditional expectation of the sine and the conditional expectation of the cosine of the response variable. Therefore, regression estimators or goodness-of-fit tests must be specifically designed and analyzed. Although for a single covariate, the problem of modeling and analyzing linear-circular regression models was considered using parametric (Fisher and Lee, 1992, 1994) and nonparametric methods (Di Marzio et al., 2012, 2013), this is not the case for several covariates, in presence of independent (or even spatial dependent) data.

The goal of this thesis is twofold. First, to analyze some inference problems for regression models with Euclidean response and covariates, and spatially correlated errors. The main contribution in this part is to propose a testing procedure for parametric regression models (with univariate response and d random covariates) in the presence of spatial correlation. The proposed test statistic is based on a comparison between a smoothed version of a parametric fit and a nonparametric estimator of the regression function, using a weighted L_2-distance. The methodology is also studied in the particular case of spatial trends with fixed design. The second aim is to adapt the approaches used to address these inference problems for more complex data, such as circular data. Considering a regression model with circular response and d real-valued covariates, nonparametric proposals to estimate the regression function are provided. The problem of estimating the circular regression function in presence of spatial correlation is also addressed. Moreover, goodness-of-fit tests for assessing parametric circular regression functions are presented for independent and spatially correlated data.

This chapter gives a brief background of some methods to address inference problems involving Euclidean and circular data. Section 1.1 is devoted to present both parametric and nonparametric methods to perform regression estimation in different frameworks. The need of proposing goodness-of-fit tests is motivated in Section 1.2. A linear-circular regression model is introduced in Section 1.3. Finally, Section 1.4 contains the outline of this dissertation.

1.1 A review on regression estimation

Let $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ be a random sample from the $(d+1)$-valued random vectors $(\mathbf{X}, Z)$, where Z denotes a scalar response which depends on a d-dimensional fixed or random covariate $\mathbf{X}$, with support $\mathcal{D} \subset \mathbb{R}^d$ (under fixed design, the covariate will be denoted by $\mathbf{x}$). For random designs, the probability density function of the design variable $\mathbf{X}$ will be denoted by f. Assume the following regression model:

$$Z_i = m(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where m is the trend or regression function, and the ε_i are random variables, which are supposed to be zero mean and second-order stationary, with covariogram or covariance function

$$C(\mathbf{X}_i - \mathbf{X}_j) = \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho_n(\mathbf{X}_i - \mathbf{X}_j), \quad i, j = 1, \ldots, n, \tag{1.2}$$

where σ^2 is the variance of the errors and ρ_n is a continuous stationary correlation function satisfying $\rho_n(\mathbf{0}) = 1$, $\rho_n(\mathbf{x}) = \rho_n(-\mathbf{x})$, and $|\rho_n(\mathbf{x})| \leq 1$, $\forall \mathbf{x} \in \mathcal{D}$. The subscript n in ρ_n allows the correlation function to shrink as $n \to \infty$ (this will be discussed more precisely in Section 1.1.2). Notice that if $C(\mathbf{X}_i - \mathbf{X}_j)$ is replaced by $C(\|\mathbf{X}_i - \mathbf{X}_j\|)$ in (1.2), where $\|\cdot\|$ represents the Euclidean norm, then the second-order stationary process is also isotropic. That is, second-order stationarity means that the dependence between two observations is a function of the difference vector between the locations where the observations are taken, while isotropy goes further, considering that the dependence is a function of the distance, ignoring the direction of the difference vector between locations. Otherwise, the process will be anisotropic. From now on, no matter if the process is isotropic or anisotropic, the covariance matrix of the errors will be denoted by $\boldsymbol{\Sigma}$, being $\boldsymbol{\Sigma}(i, j) = \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j)$, for $i, j = 1, \ldots, n$, its (i, j)-entry.

With the aim of characterizing the regression function $m(\mathbf{x}) = \mathbb{E}(Z \mid \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathcal{D}$, in (1.1), parametric regression models may be fit, allowing for a direct interpretation of the corresponding parameter values. Assuming that $m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}$, where $\boldsymbol{\mathcal{B}} \subset \mathbb{R}^q$ is a compact set, and q denotes the dimension of the parameter space $\boldsymbol{\mathcal{B}}$, then estimating $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, a parametric regression estimator of $m_{\boldsymbol{\beta}}$, denoted by $m_{\hat{\boldsymbol{\beta}}}$, is obtained. Assuming only some regularity conditions, nonparametric fits also provide a global view of the mean of the process. Their flexibility allows to model complex relations beyond a parametric form. In this section, a review on regression estimation for multivariate data is provided. Existing developments for spatially correlated data are also presented.

In what follows, $g^{(p)}(x)$ will denote the p-order derivative of a sufficiently smooth real function g at x. Moreover, if g is a sufficiently smooth real multivariate function, $\boldsymbol{\nabla} g(\mathbf{x})$ and $\boldsymbol{\mathcal{H}}_g(\mathbf{x})$ will denote the vector of first-order partial derivatives and the Hessian matrix of g at $\mathbf{x}$, respectively. For a vector $\mathbf{u} = (u_1, \ldots, u_d)^{\mathrm{T}}$ and an

integrable function g, the multiple integral $\int\int\cdots\int g(\mathbf{u})du_1du_2\ldots du_d$ will be simply denoted as $\int g(\mathbf{u})d\mathbf{u}$. We use $\mathbf{1}_d$ and $\mathbf{1}_{d\times d}$ to denote a $d\times 1$ vector and a $d\times d$ matrix with every entry equal to 1, respectively. In addition, $\mathbf{I}_d$ denotes the $d\times d$ identity matrix. Finally, for any matrix $\mathbf{A}$, $\mathbf{A}^{\mathrm{T}}$, $|\mathbf{A}|$, $\mathrm{tr}(\mathbf{A})$, $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ denote its transpose, determinant, trace, maximum eigenvalue and minimum eigenvalue, respectively.

1.1.1 Regression estimation for multivariate data

Assume regression model (1.1) holds, but with independent errors. As pointed out in the previous section, to estimate the regression function, parametric and nonparametric methods can be employed. Considering that the regression function belongs to a specific parametric family, standard procedures based on least squares or maximum likelihood could be used to obtain a parametric estimator of m (see Rao, 1973). Assuming that $m\in\mathcal{M}_{\boldsymbol{\beta}}$, the least squares regression estimator of $\boldsymbol{\beta}$ is obtained by solving the minimization problem:

$$\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}\left[Z_i-m_{\boldsymbol{\beta}}(\mathbf{X}_i)\right]^2. \tag{1.3}$$

The multiple linear regression model is a parametric model in which the regression function is linear in the parameters. If $X_1,\ldots,X_d$ are the predictor variables and $\boldsymbol{\beta}=(\beta_0,\ldots,\beta_d)$, then $m_{\boldsymbol{\beta}}(\mathbf{X})=\beta_0+\beta_1X_1+\cdots+\beta_dX_d$. The main advantage of this model is its simplicity, however, it is rather restrictive. Moreover, for inference purposes, some regularity assumptions (apart from linearity), such as normality, independence and homoscedasticity are usually assumed by the regression errors. For linear regression models, the least squares problem given in (1.3) is expressed as:

$$\min_{\boldsymbol{\beta}}\sum_{i=1}^{n}\left(Z_i-\beta_0-\sum_{j=1}^{d}\beta_jX_{ij}\right)^2, \tag{1.4}$$

where X_{ij} denotes the ith observation of the jth predictor variable. Other estimation procedures, such as maximum likelihood could be used to estimate parametrically the regression function. This method consists in maximizing a likelihood function,

sometimes involving some problems in the optimization. Under the assumption of normality, both least squares and maximum likelihood approaches provide the same estimate of $\boldsymbol{\beta}$.

Despite parametric models lead to optimal results (in terms of efficiency) if the parametric assumption holds, conclusions can be totally misleading if this assumption fails. Nonparametric regression estimation is appealing since only some regularity conditions must be assumed. In this research, we will focus on kernel smoothing methods. Other nonparametric procedures include splines, wavelets or orthogonal series methods, but they are out of the scope of this thesis. Given that $m(\mathbf{x}) = \mathbb{E}(Z \mid \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathcal{D}$, estimators of m are generally defined as a locally weighted average of the response variable or from local polynomial regression. Both approaches are described below.

Locally weighted average estimators

A nonparametric estimator of m can be obtained computing a locally weighted average of the response variable:

$$\hat{m}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} W_{\mathbf{H},i}(\mathbf{x}) Z_i, \tag{1.5}$$

where $W_{\mathbf{H},i}$ denotes a weight function, depending on a smoothing parameter which controls the neighborhood of each observation, and must be selected. An example of such an estimator is the Nadaraya–Watson estimator (see Nadaraya, 1964; Watson, 1964). The Nadaraya–Watson estimator can be seen as a particular case of the wider class of nonparametric estimators, the so-called local polynomial estimators (of degree p), which are constructed by solving a weighted least squares problem involving polynomials (of degree p). The Nadaraya–Watson estimator is obtained when the degree of the polynomial is equal to zero (local constant). For this reason, denoting by $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$ the local polynomial estimator of degree p, $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ will denote the Nadaraya–Watson estimator. For the Nadaraya–Watson estimator, the weights in expression (1.5) are $W_{\mathbf{H},i}(\mathbf{x}) = K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})/[1/n \sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})]$, where for $\mathbf{u} \in \mathcal{D}$, $K_{\mathbf{H}}(\mathbf{u}) = |\mathbf{H}|^{-1} K(\mathbf{H}^{-1}\mathbf{u})$ is the rescaled version of a multivariate kernel function K and $\mathbf{H}$ is a $d \times d$ symmetric positive definite matrix. Therefore, the

Kernel	Function
Uniform	$K(u) = \frac{1}{2}\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Epanechnikov	$K(u) = \frac{3}{4}(1-u^2)\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Biweight	$K(u) = \frac{15}{16}(1-u^2)^2\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Triweight	$K(u) = \frac{35}{32}(1-u^2)^3\mathbb{I}_{\{\lvert u\rvert \leq 1\}}$
Gaussian	$K(u) = \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)$

Table 1.1: Common second-order univariate kernel functions.

Nadaraya–Watson or local constant estimator is defined by:

$$\hat{m}_{\mathbf{H}}(\mathbf{x};0) = \frac{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})Z_i}{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})}. \tag{1.6}$$

A common technique for generating multivariate kernels is the use of the product of univariate kernels. Denoting by $\mathbb{I}_{\{.\}}$ the indicator function, usual second-order univariate kernels (their first moment is zero and the second one is finite) are given in Table 1.1 and plotted in Figure 1.1. Two examples of d-dimensional kernels, with non-zero value if $\|\mathbf{u}\| < 1$, are the uniform kernel

$$K(\mathbf{u}) = 1/V_d,$$

and the Epanechnikov kernel

$$K(\mathbf{u}) = \frac{d(d+2)}{2S_d}(1 - \|\mathbf{u}\|^2)\mathbb{I}_{\{\|\mathbf{u}\|<1\}},$$

where V_d and S_d are the volume and the area of the surface of the unit sphere in $\mathbb{R}^d$, respectively. In the Nadaraya–Watson estimator given in (1.6), the smoothing or bandwidth matrix $\mathbf{H}$ controls the shape and the size of the local neighborhood used to estimate m and its selection plays an important role in the estimation process. The effect of the bandwidth when $d = 1$ is analyzed in a simulated dataset. A sample of size $n = 100$ is generated from a regression model with explanatory variable drawn from a $N(0, \sigma)$, with standard deviation $\sigma = 2$, regression function $m(x) = x^2 + \sin(x)$, and random errors following a $N(0, 2)$. Denoting by h the smoothing parameter when $d = 1$, Figure 1.2 shows the regression function (black line) and the

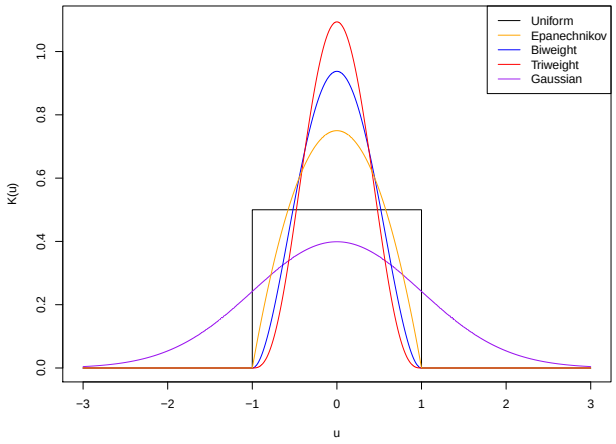

Figure 1.1: Some univariate kernel functions: uniform (black line), Epanechnikov (orange line), biweight (blue line), triweight (red line) and Gaussian (purple line).

Nadaraya–Watson estimator with Gaussian kernel and $h = 0.1$ (green line), $h = 0.4$ (red line) and $h = 1$ (blue line). If the bandwidth parameter is small, the number of observations effectively accounted for the regression function estimation will be small, and an undersmoothed curve will be obtained (green line). Conversely, if the smoothing parameter is large, too many observations will be considered to estimate the regression function at every point, leading to oversmoothing (blue line).

The asymptotic conditional bias and variance of the Nadaraya–Watson estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ were derived by Härdle and Müller (2012). The following assumptions on the design and on the nonparametric estimator of the regression functions are needed:

(A1) The design density f is continuously differentiable at $\mathbf{x} \in \mathcal{D}$, and satisfies $f(\mathbf{x}) > 0$.

(A2) All second-order derivatives of the regression functions m are continuous at $\mathbf{x}$.

(H1) The bandwidth matrix $\mathbf{H}$ is symmetric and positive definite, with $\mathbf{H} \to \mathbf{0}$ and $n|\mathbf{H}| \to \infty$, as $n \to \infty$.

(K1) The kernel K is bounded with compact support (for simplicity with a non-zero value only if $\|\mathbf{u}\| \leq 1$) and satisfies $\int K(\mathbf{u})d\mathbf{u} = 1$, $\mathbf{u} \in \mathcal{D}$. Moreover, all odd-order moments of K vanish. It is also assumed that $R(K) = \int K^2(\mathbf{u})d\mathbf{u} < \infty$.

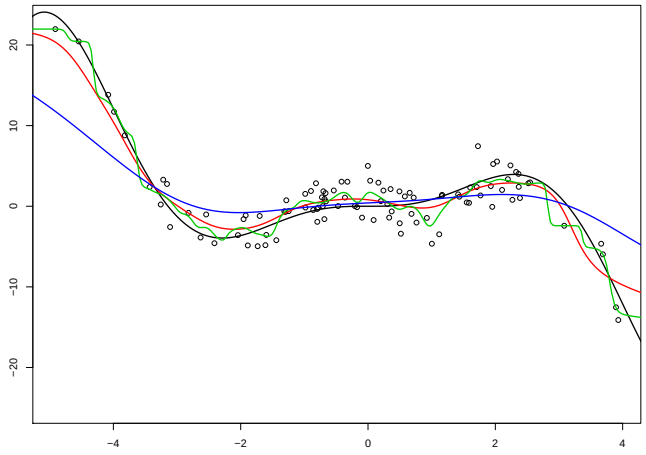

Figure 1.2: Nadaraya–Watson estimator with Gaussian kernel and $h = 0.1$ (green line), $h = 0.4$ (red line) and $h = 1$ (blue line). Sample of size $n = 100$ generated from a regression model with $X \sim N(0, \sigma)$, with standard deviation $\sigma = 2$, $m(x) = x^2 + \sin(x)$ (black line), and $\varepsilon \sim N(0, 2)$.

In assumption (H1), $\mathbf{H} \to \mathbf{0}$ means that every entry of $\mathbf{H}$ goes to $\mathbf{0}$. Notice that, since $\mathbf{H}$ is symmetric and positive definite, $\mathbf{H} \to \mathbf{0}$ is equivalent to $\lambda_{\max}(\mathbf{H}) \to 0$. $|\mathbf{H}|$ is a quantity of order $\mathcal{O}\left[\lambda_{\max}^d(\mathbf{H})\right]$ since $|\mathbf{H}|$ is equal to the product of all eigenvalues of $\mathbf{H}$. Denoting by $\int \mathbf{u}\mathbf{u}^{\mathrm{T}} K(\mathbf{u}) d\mathbf{u} = \mu_2(K)\mathbf{I}_d$, where $\mu_2(K) \neq 0$ (the values of μ_2 for different spherically symmetric kernels and for different dimensions d are provided by Duong, 2015), and assuming (A1), (A2), (H1) and (K1), the asymptotic conditional bias of estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, at a point $\mathbf{x}$ in the interior of the support of f, is:

$$\begin{aligned} \mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 0) - m(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_m(\mathbf{x})] \\ &\quad + \frac{\mu_2(K)}{f(\mathbf{x})} \boldsymbol{\nabla}^{\mathrm{T}} m(\mathbf{x}) \mathbf{H}^2 \boldsymbol{\nabla} f(\mathbf{x}) \\ &\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \end{aligned} \tag{1.7}$$

and the asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{R(K)}{n|\mathbf{H}| f(\mathbf{x})}\sigma^2 + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \tag{1.8}$$

The Nadaraya–Watson estimator has a large bias at the region where the gradi-

ent of the regression function or the ratio $\boldsymbol{\nabla} f(\mathbf{x})/f(\mathbf{x})$ is large (see, Fan and Gijbels, 1996). Moreover, its asymptotic properties are complicated to derive because its random denominator. Initially, this motivates the introduction of other local weighted average estimators such as the Gasser–Müller (Gasser and Müller, 1979) or the Priestley–Chao (Priestley and Chao, 1972) estimator. The Gasser–Müller estimator reduces the bias of the Nadaraya–Watson estimator, but at the expense of increasing its variance. Moreover, this estimator is difficult to extend to a higher dimensional setting, because its expression involve sorting design points, which is not a computationally trivial problem in higher dimensional spaces.

Local polynomial estimators

The motivation for the local polynomial fit comes from attempting to find an estimator of the regression function m, minimizing the residual sum of squares without assuming any particular form of m. The idea of local polynomial regression was introduced by Stone (1977) and Cleveland (1979). It was studied by Fan (1992), Fan and Gijbels (1992), Fan and Gijbels (1995) and Fan et al. (1997), among others. Fan et al. (1993) showed that local polynomial regression estimators have advantages over local weighted average estimators in terms of design adaptation and high asymptotic efficiency.

As it was pointed out before, the Nadaraya–Watson estimator can be obtained from local polynomial regression, when the degree of the polynomial to fit is zero. The local linear estimator (polynomial degree one) at a given point $\mathbf{x}$ is the solution for α_1 to the least squares minimization problem

$$\min_{\alpha_1,\boldsymbol{\beta}_1} \sum_{i=1}^{n} \left[Z_i - \alpha_1 - \boldsymbol{\beta}_1^{\mathrm{T}}(\mathbf{X}_i - \mathbf{x})\right]^2 K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}).$$

The local linear estimator can be explicitly written as:

$$\hat{m}_{\mathbf{H}}(\mathbf{x}; 1) = \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{Z}, \tag{1.9}$$

where $\mathbf{e}_1$ is a $(d+1)$ vector with 1 in the first entry and all other entries 0, $\boldsymbol{\mathcal{X}}_{\mathbf{x}}$ is a matrix having $[1, (\mathbf{X}_i - \mathbf{x})^{\mathrm{T}}]$ as its ith row, $\boldsymbol{\mathcal{W}}_{\mathbf{x}} = \mathrm{diag}\{K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{x}), \ldots, K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{x})\}$

and $\mathbf{Z} = (Z_1, \ldots, Z_n)^{\mathrm{T}}$. This estimator can be expressed as a locally weighted average type estimator, considering the weights $W_{\mathbf{H},i}(\mathbf{x}) = \mathbf{e}_1^{\mathrm{T}} \left(n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}}\right)^{-1} [1, (\mathbf{X}_i - \mathbf{x})^{\mathrm{T}}] K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})$ in equation (1.5).

Under assumptions (A1), (A2), (H1) and (K1), asymptotic properties of the local linear estimator were derived by Ruppert and Wand (1994). The asymptotic conditional bias of estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, at a point $\mathbf{x}$ in the interior of the support of f, is:

$$\mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 1) - m(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_m(\mathbf{x})] + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \quad (1.10)$$

and the asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 1) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{R(K)}{n|\mathbf{H}|f(\mathbf{x})}\sigma^2 + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \quad (1.11)$$

A comparison of the asymptotic bias and variance of Nadaraya–Watson (1.6) and local linear estimators (1.9) shows that the Nadaraya–Watson estimator presents a larger bias, in particular, in the region where the gradient of the regression function m or of the design density f is large. This estimator also has a larger bias order when estimating the regression function at boundary points, while the local linear estimator is efficient correcting that bias (see Fan and Gijbels, 1996). The local linear estimator has the same bias order in the interior as well as in the boundary of the support of f. The asymptotic conditional variance of both estimators is the same. Comparisons between Nadaraya–Watson and local linear estimators were discussed in detail by Chu et al. (1991), Fan and Gijbels (1992) and Hastie and Loader (1993).

Considering that the degree of the polynomial p is larger than one, asymptotic properties of the local polynomial estimator were also studied by Ruppert and Wand (1994). The authors derived the close expression of the local quadratic estimator, as well as its asymptotic bias and variance. The conditional bias of local polynomials of degree p will be of order $\mathcal{O}_{\mathbb{P}}\{[\mathrm{tr}(\mathbf{H}^2)]^{(p+1)/2}\}$. Moreover, if p is even and f has a continuous derivative in a neighborhood of $\mathbf{x}$, being $\mathbf{x}$ an interior point of the support of the design density f, then the bias will be of order $\mathcal{O}_{\mathbb{P}}\{[\mathrm{tr}(\mathbf{H}^2)^{p/2+1}]\}$. For simplicity and given that for any dimension d the generalization to higher polynomial degree involves complicated expressions, the authors derived asymptotic properties

of the local polynomial estimator for a general degree p and $d = 1$.

For the unidimensional case, assuming that the derivative of order $(p+1)$ of the regression function m exists, and using a Taylor expansion, the regression function m can be locally approximated by a polynomial of degree p,

$$m(z) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(z-x)^j = \sum_{j=0}^{p} \beta_j (z-x)^j,$$

being z a point in a neighborhood of x. The terms of the previous approximation can be estimated by solving the minimization problem:

$$\min_{\{\beta_j\}_{j=0}^{p}} \sum_{i=1}^{n} \left[Z_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j \right]^2 K_h(X_i - x), \tag{1.12}$$

where $K_h(u) = 1/hK(u/h)$, being K a univariate kernel function, and h the bandwidth or smoothing parameter. Denoting by $\hat{\beta}_j$, for $j = 0, \ldots, p$, the solutions of (1.12), the pth order local polynomial regression estimator of m at $x \in \mathbb{R}$ is $\hat{m}_h(x; p) = \hat{\beta}_0$, and it can be explicitly written as:

$$\hat{m}_h(x; p) = \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_x \boldsymbol{\mathcal{X}}_{x,p})^{-1} \boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_x \mathbf{Z}, \tag{1.13}$$

where, in this case, $\mathbf{e}_1$ denotes a $(p+1) \times 1$ vector having 1 in the first entry and zero elsewhere, $\boldsymbol{\mathcal{X}}_{x,p}$ is a $n \times p$ matrix with the (i,k)-entry equal to $(X_i - x)^{k-1}$, and $\boldsymbol{\mathcal{W}}_x$ is a diagonal matrix of order n with the c(i,i)-entry equal to $K_h(X_i - x)$.

Notice that from (1.12), estimates of the derivatives of the regression function m could be also obtained. The least squares problem given in (1.4) when $d = 1$ can be seen as a *global* version of the locally weighted least squares problem given in (1.12). Simple linear regression corresponds to local polynomial regression if the degree of the polynomial is equal to one, the weight function is constant and $h = \infty$.

The asymptotic properties of the local polynomial estimator, for a general p, given in (1.13), were derived by Ruppert and Wand (1994). Let $K_{(p)}$ be the equivalent kernel function defined in Lejeune and Sarda (1992), which is a kernel of order $(p+2)$ when p is even and of order $(p+1)$ otherwise. Let $\mu_j(K_{(p)})$ and $R(K_{(p)})$ denote the moment of order j and the "roughness" of $K_{(p)}$, respectively. Moreover, assume that

the design density f is continuously differentiable at $x \in \mathcal{D} \subset \mathbb{R}$, with $f(x) > 0$, the bandwidth h satisfies $h \to 0$ and $nh \to \infty$, as $n \to \infty$, the kernel K is a symmetric density function, twice continuously differentiable and with compact support, and that m admits continuous derivatives up to order $(p+2)$ in a neighborhood of x. If x is an interior point of the support of the design density f, then, for even p,

$$\begin{aligned}\mathbb{E}[\hat{m}_h(x;p) - m(x) \mid X_1, \ldots, X_n] &= \frac{h^{p+2}\mu_{p+2}(K_{(p)})f^{(1)}(x)}{(p+1)!f(x)}m^{(p+1)}(x) \\ &+ \frac{h^{p+2}\mu_{p+2}(K_{(p)})}{(p+2)!}m^{(p+2)}(x) \\ &+ o_{\mathbb{P}}\left(h^{p+2}\right), \end{aligned} \tag{1.14}$$

while for odd p,

$$\mathbb{E}[\hat{m}_h(x;p) - m(x) \mid X_1, \ldots, X_n] = \frac{h^{p+1}\mu_{p+1}(K_{(p)})}{(p+1)!}m^{(p+1)}(x) + o_{\mathbb{P}}\left(h^{p+1}\right), \tag{1.15}$$

and in both cases,

$$\mathbb{V}\mathrm{ar}[\hat{m}_h(x;p) \mid X_1, \ldots, X_n] = \frac{R\left(K_{(p)}\right)}{nhf(x)}\sigma^2 + o_{\mathbb{P}}\left(\frac{1}{nh}\right). \tag{1.16}$$

An important issue in local polynomial regression is the order of the polynomial to be fit. Higher-order polynomials allow a precise fitting, leading to a possible bias reduction, but with an increase in the variance, due to introducing more parameters. The asymptotic variance of $\hat{m}_h(x;p)$ only increases whenever p goes from an odd order to the following even order. For example, there is no difference when going from $p = 0$ to $p = 1$, but when going from $p = 1$ to $p = 2$, the asymptotic variance increases, because $R(K_{(0)})$ is equal to $R(K_{(1)})$, but $R(K_{(2)})$ is larger than $R(K_{(1)})$. Odd orders are preferred, since the gain in bias does not lead to an increase in variance. In the case of the regression function, Fan and Gijbels (1996) recommended to use polynomial orders $p = 1$ or $p = 3$ for estimating this curve.

The use of local polynomial methods to estimate the regression function involves a bandwidth selection. For d covariates, a $d \times d$ bandwidth matrix must be properly chosen to avoid producing an undersmoothed estimator (with high variability) or an

oversmoothed estimator (probably, with larger bias). See Figure 1.2 to observe the effect of the bandwidth when using the Nadaraya–Watson estimator and $d = 1$. This issue will be discussed below.

Bandwidth matrix selection

In any kernel curve estimation problem, the choice of the smoothing parameter is crucial and, hence, this should be done with extreme care. For kernel-type regression estimators, several bandwidth selection methods have been proposed in the literature. An optimal local bandwidth (which changes for each $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$) for the Nadaraya–Watson estimator, $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, can be obtained by minimizing the (conditional) mean squared error (MSE), given by:

$$\mathrm{MSE}[\hat{m}_{\mathbf{H}}(\mathbf{x};0)] = \{\mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x};0) - m(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n]\}^2 + \mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n].$$

The optimal local bandwidth can be approximated by its asymptotic version, obtained by minimizing the asymptotically mean squared error (AMSE), which is defined using the leading terms of the (conditional) asymptotic bias and variance of $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$ given in (1.7) and (1.8), by:

$$\begin{aligned}
\mathrm{AMSE}[\hat{m}_{\mathbf{H}}(\mathbf{x};0)] &= \left\{ \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_m(\mathbf{x})] + \frac{\mu_2(K)}{f(\mathbf{x})} \boldsymbol{\nabla}^{\mathrm{T}} m(\mathbf{x}) \mathbf{H}^2 \boldsymbol{\nabla} f(\mathbf{x}) \right\}^2 \\
&\quad + \frac{R(K)\sigma^2}{n|\mathbf{H}|f(\mathbf{x})} \\
&= \frac{1}{4}\mu_2^2(K)\mathrm{tr}^2 \left\{ \mathbf{H}^2 \Big[\frac{\boldsymbol{\nabla} f(\mathbf{x}) \boldsymbol{\nabla}^{\mathrm{T}} m(\mathbf{x})}{f(\mathbf{x})} + \frac{\boldsymbol{\nabla} m(\mathbf{x}) \boldsymbol{\nabla}^{\mathrm{T}} f(\mathbf{x})}{f(\mathbf{x})} \right. \\
&\quad \left. + \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \Big] \right\} + \frac{R(K)\sigma^2}{n|\mathbf{H}|f(\mathbf{x})}. \qquad (1.17)
\end{aligned}$$

The minimizer of equation (1.17), with respect to $\mathbf{H}$, provides an asymptotically optimal local bandwidth matrix for $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, which is given by:

$$\mathbf{H}_{\mathrm{opt}}(\mathbf{x};0) = \left[\frac{R(K)\sigma^2}{nd\mu_2^2(K) f(\mathbf{x})} |\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x})|^{1/2} \right]^{1/d+4} \cdot \left[\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x}) \right]^{-1/2}, \qquad (1.18)$$

where

$$\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x}) = \begin{cases} \boldsymbol{\mathcal{G}}(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{G}}(\mathbf{x}) \text{ is positive definite,} \\ -\boldsymbol{\mathcal{G}}(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{G}}(\mathbf{x}) \text{ is negative definite,} \end{cases}$$

with

$$\boldsymbol{\mathcal{G}}(\mathbf{x}) = \frac{\boldsymbol{\nabla} f(\mathbf{x}) \boldsymbol{\nabla}^{\mathrm{T}} m(\mathbf{x})}{f(\mathbf{x})} + \frac{\boldsymbol{\nabla} m(\mathbf{x}) \boldsymbol{\nabla}^{\mathrm{T}} f(\mathbf{x})}{f(\mathbf{x})} + \boldsymbol{\mathcal{H}}_m(\mathbf{x}).$$

This optimization result can be proved using Proposition A.3 in Appendix A. Note that in the expression of $\mathbf{H}_{\mathrm{opt}}(\mathbf{x}; 0)$, the matrix $[\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x})]^{-1/2}$ determines the shape and the orientation in the d-dimensional space of the covariate region which is used to locally compute the estimator. Such data regions for computing the estimator are ellipsoids in $\mathbb{R}^d$, being the magnitude of the axes controlled by $\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x})$. In the particular case of $\mathbf{H} = h\mathbf{I}_d$, the estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, with $\mathbf{x}$ being an interior point of the support, achieves an optimal convergence rate of $n^{-4/(d+4)}$. As a consequence of (1.10) and (1.11), and similarly to the Nadaraya–Watson case, an asymptotically optimal local bandwidth can be also obtained for $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, and it is given by:

$$\mathbf{H}_{\mathrm{opt}}(\mathbf{x}; 1) \quad = \quad \left[\frac{R(K)\sigma^2}{nd\mu_2^2(K) f(\mathbf{x})} |\tilde{\boldsymbol{\mathcal{H}}}_m(\mathbf{x})|^{1/2} \right]^{1/d+4} \cdot \left[\tilde{\boldsymbol{\mathcal{H}}}_m(\mathbf{x}) \right]^{-1/2}, \qquad (1.19)$$

where

$$\tilde{\boldsymbol{\mathcal{H}}}_m(\mathbf{x}) = \begin{cases} \boldsymbol{\mathcal{H}}_m(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \text{ is positive definite,} \\ -\boldsymbol{\mathcal{H}}_m(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{H}}_m(\mathbf{x}) \text{ is negative definite.} \end{cases}$$

In the univariate case, a local bandwidth for the local polynomial estimator $\hat{m}_h(x; p)$ was discussed in Fan et al. (1996). The case where p is even involves a more complicated approximation in the bias term, so we only consider the case where p is odd. From (1.15) and (1.16), it can be obtained that the minimization of the AMSE leads to

$$h_{\mathrm{opt}}(x; p) = C_p(K) \left\{ \frac{\sigma^2}{[m^{(p+1)}(x)]^2 f(x)} \right\}^{1/(2p+3)} n^{-1/(2p+3)}, \qquad (1.20)$$

where

$$C_p(K) = \left\{ \frac{(p+1)!^2 \int K_{(p)}^2(t) dt}{2(p+1)[\int t^{p+1} K_{(p)}(t) dt]^2} \right\}^{1/(2p+3)}.$$

The use of the local optimal bandwidths given in (1.18), (1.19) and (1.20), is lim-

ited in practice, since they depend on unknown functions, such as the design density f, the variance σ^2, the Hessian matrix of an unknown regression function m in the case of (1.18) and (1.19), and the $(p+1)$-derivative of m, in the case of (1.20). In addition, when the goal is to reconstruct the whole regression function and the focus is not only set on a specific point, it is more usual in practice to consider a global bandwidth for estimation rather than pursuing an estimator based on local bandwidths. For the univariate case, an optimal global bandwidth, which was derived by Fan and Gijbels (1992), could be obtained by minimizing the (conditional) weighted mean integrated squared error (MISE):

$$\begin{aligned}\mathrm{MISE}[\hat{m}_h(x;p)] &= \int (\{\mathbb{E}[\hat{m}_h(x;p) - m(x) \mid X_1, \ldots, X_n]\}^2 \\ &\quad + \mathbb{V}\mathrm{ar}[\hat{m}_h(x;p) \mid X_1, \ldots, X_n]) w(x) dx,\end{aligned}$$

being w a weight function. Using (1.15) and (1.16), it can be obtained that the minimization of the (conditional) asymptotic MISE (AMISE) leads to the global bandwidth

$$h^p_{\mathrm{opt}} = C_p(K) \left\{ \frac{\sigma^2 \int w(x)/f(x) dx}{\int [m^{(p+1)}(x)]^2 w(x) dx} \right\}^{1/(2p+3)} n^{-1/(2p+3)}. \tag{1.21}$$

The resulting bandwidth also depends on unknown quantities, which must be estimated in practice. Using the rule-of-thumb method, the approach would consist in fitting a polynomial of order $(p+3)$ globally to m, leading to the parametric fit $\check{m}$. Denoting by $\check{\sigma}^2$ the standardized residual sum of squares from this parametric fit and by $\check{m}^{(p+1)}$ the $(p+1)$-derivative function of $\check{m}$, and taking $w(x) = f(x) w_0(x)$, for some specific function w_0, from (1.21), the rule-of-thumb bandwidth selector is:

$$h^p_{\mathrm{rot}} = C_p(K) \left\{ \frac{\check{\sigma}^2 \int w_0(x) dx}{\sum_{i=1}^n [\check{m}^{(p+1)}(X_i)]^2 w_0(X_i)} \right\}^{1/(2p+3)} n^{-1/(2p+3)}.$$

This bandwidth provides an initial guess for the amount of smoothing. However, more carefully procedures to estimate the unknown quantities in (1.21) are usually employed, leading to plug-in bandwidth selectors. The ideas of plug-in bandwidth selection to develop strategies for choosing the smoothing parameter were employed

by Ruppert et al. (1995). The authors proposed three plug-in type selectors to obtain the smoothing parameter, minimizing an asymptotic approximation to the MISE of the local polynomial estimator of the regression functions and its derivatives.

For $d > 1$, an asymptotic global optimal bandwidth matrix $\mathbf{H}$ could be also obtained by minimizing a global error measurement. However, unfortunately, this optimization problem is not trivial, not being possible to obtain a closed form solution. Alternatively, a cross-validation (CV) method can be used to select a bandwidth matrix (Bowman, 1984; Cleveland, 1979). In this case, for a general dimension d, the smoothing parameter is chosen selecting the bandwidth matrix $\mathbf{H}$ which minimizes the function:

$$\mathrm{CV}(\mathbf{H}) = \sum_{i=1}^{n} [Z_i - \hat{m}_{\mathbf{H},-i}(\mathbf{X}_i; p)]^2, \tag{1.22}$$

where $\hat{m}_{\mathbf{H},-i}(\mathbf{X}_i; p)$ stands for the Nadaraya–Watson or the local linear estimators (depending on whether $p = 0$ or $p = 1$, respectively), computed using all observations except $(\mathbf{X}_i, Z_i)$ and evaluated at $\mathbf{X}_i$.

An alternative expression of (1.22) can be derived. Firstly, it holds that

$$\hat{\mathbf{m}}_{\mathbf{H};p} = \mathbf{S}_p^{\mathrm{T}} \mathbf{Z}, \quad p = 0, 1,$$

being $\hat{\mathbf{m}}_{\mathbf{H};p} = [\hat{m}_{\mathbf{H}}(\mathbf{X}_1; p), \ldots, \hat{m}_{\mathbf{H}}(\mathbf{X}_n; p)]$ and $\mathbf{S}_p$ a $n \times n$ matrix whose ith row is the vector $\mathbf{S}_{\mathbf{X_i};p}$, for $p = 0, 1$, with

$$\mathbf{S}_{\mathbf{X_i};0} = \left[\frac{K_{\mathbf{H}}(\mathbf{X}_1 - \mathbf{X}_i)}{\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{X}_i)}, \ldots, \frac{K_{\mathbf{H}}(\mathbf{X}_n - \mathbf{X}_i)}{\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{X}_i)} \right], \tag{1.23}$$

and

$$S_{\mathbf{X_i};1} = \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{\mathbf{X}_i}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{X}_i} \boldsymbol{\mathcal{X}}_{\mathbf{X}_i})^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{X}_i}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{X}_i}, \tag{1.24}$$

where $\boldsymbol{\mathcal{X}}_{\mathbf{X}_i}$ and $\boldsymbol{\mathcal{W}}_{\mathbf{X}_i}$ were defined in equation (1.9), for $\mathbf{x} = \mathbf{X}_i$. The smoothing

matrix $\mathbf{S}_p$, for $p = 0, 1$, satisfies $\mathbf{S}_p\mathbf{1}_n = \mathbf{1}_n$ and, therefore,

$$\hat{m}_{\mathbf{H},-i}(\mathbf{X}_i; p) = \frac{\hat{m}_{\mathbf{H}}(\mathbf{X}_i; p) - s_{ii,p}Z_i}{1 - s_{ii,p}},$$

where $s_{ii,p}$ is the i-entry of the smoothing vector $\mathbf{S}_{\mathbf{X_i};p}$.

Consequently, the expression given in (1.22) can be written as follows:

$$\mathrm{CV}(\mathbf{H}) = \sum_{i=1}^{n} \left[\frac{Z_i - \hat{m}_{\mathbf{H}}(\mathbf{X}_i; p)}{1 - s_{ii,p}} \right]^2.$$

The previous expression has the potential to be computationally less expensive to be implemented than (1.22), since the model does not have to be fit n times. Notice that (1.22) can be very time consuming if n is large, and if each individual model is slow to fit. Other criteria, such as the generalized cross-validation (GCV) (Craven and Wahba, 1979) are often used. GCV is defined as CV, but replacing the diagonal terms $s_{ii,p}$ by the average diagonal term $\mathrm{tr}(\mathbf{S}_p)$. For GCV, the smoothing parameter is chosen by selecting the matrix bandwidth $\mathbf{H}$ minimizing the function

$$\mathrm{GCV}(\mathbf{H}) = \sum_{i=1}^{n} \left[\frac{Z_i - \hat{m}_{\mathbf{H}}(\mathbf{X}_i; p)}{1 - \frac{1}{n}\mathrm{tr}(\mathbf{S}_p)} \right]^2.$$

The GCV method can be regarded as an approximation to CV. One of the motivations of using GCV instead of CV is to save on calculations, because GCV does not require to know each of the elements $s_{ii,p}$ of the smoother matrix $\mathbf{S}_p$. For GCV, the individual elements $s_{ii,p}$ are replaced by their average value, which is obtained by calculating the trace of the smoother matrix.

1.1.2 Regression estimation for spatially correlated data

In this section, multivariate regression estimation for spatially correlated data is presented. Considering regression model (1.1) holds, a brief review of parametric and nonparametric regression estimation for spatially correlated data is given. In order to define a parametric estimator, when dealing with these type of data, in contrast to independent data, stationarity conditions about the error process in model (1.1) are

assumed. For the nonparametric approach, kernel-type estimators defined in Section 1.1.1 will be studied in this framework.

Parametric regression estimation for spatially correlated data

When the data are spatially correlated, in order to perform statistical inference using a parametric approach, it is usual to introduce conditions about the error in model (1.1), assuming stationarity of some kind. In particular, for a proper regression estimation, the dependence structure (although not being of primary interest) must be accounted for, usually through iterative least squares procedures or maximum likelihood approaches, under stationary assumptions (see, for instance, Cressie, 1993; Diggle et al., 2010).

If second-order or intrinsic stationarity is assumed, then the dependence structure will be specified by the covariogram given in (1.2) or by the variogram $2\gamma_n$, respectively (Cressie, 1993). Second-order stationarity means that dependence between two observations is only a function of the difference vector between the locations where the observations are taken, while intrinsic stationarity implies that the variance of the difference of two observations is the same for any pair of locations whose difference is the same, since $2\gamma_n(\mathbf{X}_i - \mathbf{X}_j) = \mathbb{V}\mathrm{ar}(\varepsilon_i - \varepsilon_j \mid \mathbf{X}_i, \mathbf{X}_j)$, for $i, j = 1, \ldots, n$. Notice that second-order stationarity implies intrinsic stationarity, therefore, if second-order stationarity holds, the dependence structure could be also characterized through the variogram function, which satisfies

$$\gamma_n(\mathbf{X}_i - \mathbf{X}_j) = \sigma^2[1 - \rho_n(\mathbf{X}_i - \mathbf{X}_j)], \quad i, j = 1, \ldots, n. \tag{1.25}$$

In this essay, an abuse of notation will be made, both *semivariogram* and *variogram* refer to the function γ_n. Moreover, for simplicity, the subscript n will be sometimes omitted. It holds that $\gamma(\mathbf{0}) = 0$, but if $\gamma(\mathbf{u}) \to c_0 \neq 0$, as $\mathbf{u} \to \mathbf{0}$, then c_0 is called the *nugget effect* (Matheron, 1962). If γ is bounded and there is $\lim_{\|\mathbf{u}\|\to\infty} \gamma(\mathbf{u})$, this limit is called the *sill*. When the semivariogram has nugget effect, the difference $c_1 = \sigma^2 - c_0$ is called *partial sill*. If σ^2 is the sill, the *range* (if it exists) is a real value r such that if $\|\mathbf{u}\| \geq r$, then $\gamma(\mathbf{u}) = \sigma^2$. When second-order stationarity holds, then the *asymptotic range* can be defined as a real value r' such that if $\|\mathbf{u}\| \geq r'$, then $\gamma(\mathbf{u}) \geq c_0 + 0.95(\sigma^2 - c_0)$ (see Chiles and Delfiner, 2009). Figure 1.3

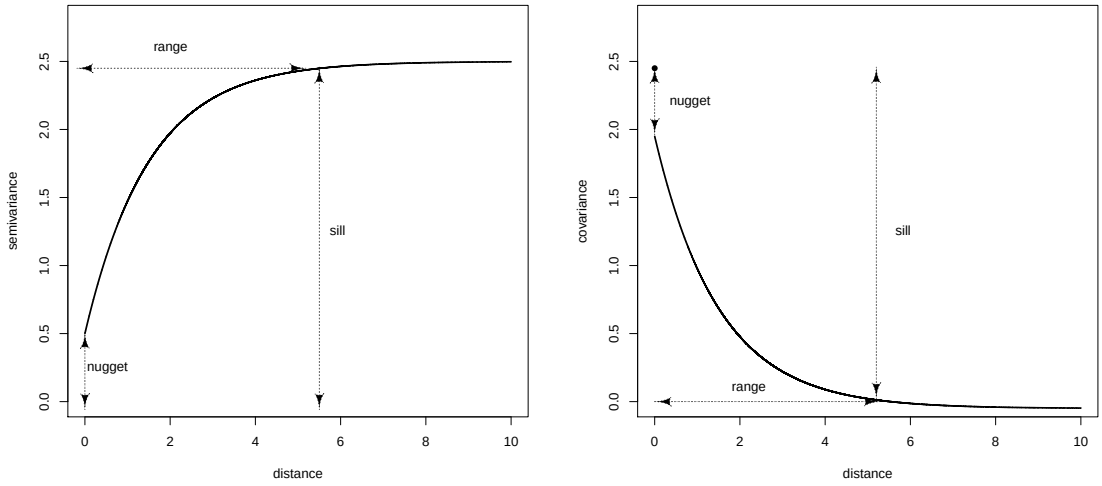

Figure 1.3: A generic variogram (left) and its corresponding covariogram (right) showing the nugget effect, the sill and the range.

shows a generic variogram (left) and its corresponding covariogram (right) identifying the role of the parameters. It should be noted that the expressions given in (1.2) and (1.25) for the covariogram and variogram functions, respectively, are correct if the nugget effect, is equal to zero. If $c_0 \neq 0$, then $\mathrm{C}(\mathbf{X}_i - \mathbf{X}_j) = c_1\rho_n(\mathbf{X}_i - \mathbf{X}_j)$, if $i \neq j$, and $\mathbb{V}\mathrm{ar}(\varepsilon) = \sigma^2$, and $\gamma_n(\mathbf{X}_i - \mathbf{X}_j) = c_0 + c_1[1 - \rho_n(\mathbf{X}_i - \mathbf{X}_j)]$, if $i \neq j$.

Considering that the regression function belongs to a parametric family, $m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}$, and that the data are spatially correlated, in order to estimate $m_{\boldsymbol{\beta}}$ accounting for the dependence structure of the error (which is also unknown, but supposing to belong to a certain parametric family), an iterative least squares procedure can be used (see Neuman and Jacobson, 1984). This method is a generalization of ordinary least squares, which takes the correlation of the data into account. Denote by $\mathbf{m}_{\boldsymbol{\beta}} = [m_{\boldsymbol{\beta}}(\mathbf{X}_1), \ldots, m_{\boldsymbol{\beta}}(\mathbf{X}_n)]^{\mathrm{T}}$, where $\mathbf{m}_{\boldsymbol{\beta}}$ collects the trend values at the observation locations under a certain parametric trend model with parameter vector $\boldsymbol{\beta}$. The specific steps of the algorithm are:

1. Obtain an initial estimator of $\boldsymbol{\beta}$ by least squares regression:

$$\tilde{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}})^{\mathrm{T}} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}}). \tag{1.26}$$

2. Using the residuals obtained with the estimation in (1.26), $\tilde{\varepsilon}_i = Z_i - m_{\tilde{\boldsymbol{\beta}}}(\mathbf{X}_i)$, $i = 1, \ldots, n$, estimate the covariance matrix of the errors, $\tilde{\boldsymbol{\Sigma}}$.

3. Update the regression parameter estimates, introducing the estimated covariance matrix $\tilde{\boldsymbol{\Sigma}}$ in the least squares minimization problem:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}})^{\mathrm{T}} \tilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{Z} - \mathbf{m}_{\boldsymbol{\beta}}). \tag{1.27}$$

   Finally, take $m_{\hat{\boldsymbol{\beta}}}$ as the parametric estimator for the regression function.

Notice that the least squares estimator $\hat{\boldsymbol{\beta}}$ given in (1.27) is a $\sqrt{n}$-consistent estimator of $\boldsymbol{\beta}$, and consequently, $m_{\hat{\boldsymbol{\beta}}}$ is a $\sqrt{n}$-consistent estimator of the parametric regression function $m_{\boldsymbol{\beta}}$.

Covariance matrix estimation in Step 2 can be carried out using different approaches. Firstly, using a parametric methodology and assuming that the variogram belongs to a valid parametric family $\{2\gamma_{\boldsymbol{\phi}},\ \boldsymbol{\phi} \in \boldsymbol{\Phi} \subset \mathbb{R}^{\vartheta}\}$ (usually $\vartheta = 3$, with the vector $\boldsymbol{\phi}$ made up of the nugget effect, the partial sill and the practical range), a parameter estimate $\hat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$ can be obtained. Following a classical approach, $\boldsymbol{\phi}$ could be approximated by fitting a parametric model to a pilot empirical variogram estimator (computed using the residuals $\tilde{\varepsilon}_i$), using a weighted least squares method (Cressie, 1985). With this parametric approximation, the variance-covariance matrix of the errors can be denoted by $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$, being $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}(i,j) = C_{\boldsymbol{\phi}}(\mathbf{X}_i - \mathbf{X}_j)$, for $i, j = 1 \ldots, n$, its (i,j)-entry. Then, replacing $\boldsymbol{\phi}$ by $\hat{\boldsymbol{\phi}}$ in these elements, a parametric estimation of $\boldsymbol{\Sigma}_{\boldsymbol{\phi}}$ could be obtained. Denoting by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\phi}}}$ this estimation, then $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}_{\hat{\boldsymbol{\phi}}}$ in Step 2 in the iterative least squares algorithm.

On the other hand, instead of using a parametric approach, flexible nonparametric variogram estimators can be employed to approximate the dependence structure. For instance, an estimate of the variogram of the residuals could be obtained as follows. First, compute a nonparametric pilot variogram estimator (Hall and Patil, 1994). A first attempt could be to use the empirical semivariogram estimator, but it may be unsatisfactory in practice (Fernández-Casal et al., 2003a). Alternatively, nonparametric kernel semivariogram estimators could be used instead, producing significantly better results than those obtained with the empirical estimator (Fernández-Casal et al., 2003b). In practice, the use of flexible models provides good approximations

to the pilot estimator avoiding misspecification problems. However, nonparametric estimators do not necessarily satisfy the conditionally negative definiteness property of a valid semivariogram. For that reason, a valid model should be fit to the nonparametric pilot estimates. For example, a flexible Shapiro–Botha variogram approach (Shapiro and Botha, 1991), fit by weighted least squares, could be employed at this step. The combination of Shapiro–Botha variogram with a nonparametric kernel semivariogram pilot estimation provides an efficient variogram estimator which can be used to estimate the corresponding covariance matrix.

Nonparametric regression estimation for spatially correlated data

When the data are correlated, the estimation of the regression function m can be also performed employing nonparametric methods. An exhaustive review on nonparametric regression estimation for correlated data, both for short-range and long-range dependence, has been provided by Opsomer et al. (2001). In the univariate case, for dependent data, asymptotic conditional bias and variance of local polynomial estimators were derived under some assumptions. For example, Masry and Fan (1997) studied this regression estimator for α-mixing and ρ-mixing time series processes. The authors obtained that the asymptotic bias and variance under dependence coincides with the result for independent observations. Francisco-Fernández and Vilar-Fernández (2001) discussed the local polynomial estimator for correlated data. A fixed regression model, assuming that the random errors present short-range dependence, was considered. This assumption is satisfied by time series that are of the form trend plus random component. The correlation is included in the asymptotic variance of the estimator. Vilar-Fernández and Francisco-Fernández (2002) proposed a nonparametric estimator of the regression function assuming that the error process follows a first-order autoregressive correlation structure. The estimator was constructed by transforming the regression model to get independent errors and then applying the local polynomial regression estimator to the new model.

For spatially correlated errors, kernel-type estimators defined in Section 1.1.1 can be employed to estimate the regression function m considering the regression model (1.1). Asymptotic properties of these estimators will depend on the behavior of the correlation function ρ_n as n increases. The asymptotic conditional bias and

variance of the Nadaraya–Watson estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, defined in (1.6), were derived by Liu (2001) under short-range dependence. In order to obtain those results, a stronger condition for the bandwidth matrix, than which was supposed in (H1) for independent data, must be assumed. Moreover, some extra hypothesis on the design and on the nonparametric estimators of the regression functions are needed:

(A3) For the correlation function ρ_n, there exist constants ρ_M and ρ_c such that $n \int |\rho_n(\mathbf{x})| d\mathbf{x} < \rho_M$ and $\lim_{n\to\infty} n \int \rho_n(\mathbf{x}) d\mathbf{x} = \rho_c$. For any sequence $\epsilon_n > 0$ satisfying $n^{1/d}\epsilon_n \to \infty$,

$$n \int_{\|\mathbf{x}\| \geq \epsilon_n} |\rho_n(\mathbf{x})| d\mathbf{x} \to 0 \quad \text{as} \quad n \to \infty.$$

(H2) The bandwidth matrix $\mathbf{H}$ is symmetric and positive definite, with $\mathbf{H} \to \mathbf{0}$ and $n|\mathbf{H}|\lambda^2_{\min}(\mathbf{H}) \to \infty$, when $n \to \infty$. The ratio $\lambda_{\max}(\mathbf{H})/\lambda_{\min}(\mathbf{H})$ is bounded above.

(K2) K is Lipschitz continuous. That is, there exists $\mathfrak{L} > 0$, such that

$$|K(\mathbf{X}_1) - K(\mathbf{X}_2)| \leq \mathfrak{L}\|\mathbf{X}_1 - \mathbf{X}_2\|, \quad \forall \mathbf{X}_1, \mathbf{X}_2 \in \mathcal{D}.$$

Assumption (A3) implies that the correlation function depends on n, and the integral $\int |\rho_n(\mathbf{x})| d\mathbf{x}$ should vanish as $n \to \infty$. The vanishing speed should not be slower than $\mathcal{O}(n^{-1})$. This assumption also implies that the integral of $|\rho_n(\mathbf{x})|$ is essentially dominated by the values of $\rho_n(\mathbf{x})$ near the origin $\mathbf{0}$. Hence, the correlation is short-range and decreases as $n \to \infty$. Arguing somewhat loosely, this can be considered as a case of increasing-domain spatial asymptotics (see Cressie, 1993), since this setup can immediately be transformed to one in which the correlation function ρ_n is fixed with respect to the sample size, but the support $\mathcal{D}$ for $\mathbf{x}$ expands (Francisco-Fernandez and Opsomer, 2005). The current setup with fixed domain $\mathcal{D}$ and shrinking ρ_n is more natural to consider when the primary purpose of the estimation is a fixed regression function m defined over a spatial domain, not the correlation function itself.

Two examples of commonly used correlation functions that satisfy the conditions

of assumption (A3) are the exponential model

$$\rho_n(\mathbf{x}) = \exp(-an\|\mathbf{x}\|), \tag{1.28}$$

and the rational quadratic model

$$\rho_n(\mathbf{x}) = \frac{1}{1 + a(n\|\mathbf{x}\|)^2}, \tag{1.29}$$

with $a > 0$ in both cases (see Cressie, 1993). In general, if $\rho_n(\mathbf{x}) = \rho(n^{1/d}\mathbf{x})$ and $\rho(\mathbf{x})$ is a fixed valid correlation function, which is continuous everywhere except at a finite number of points and absolutely integrable in $\mathbb{R}^d$, then it can be proved that $\rho_n(\mathbf{x})$ satisfies assumption (A3).

The condition $n|\mathbf{H}|\lambda^2_{\min}(\mathbf{H}) \to \infty$, when $n \to \infty$, in (H2) requires that every eigenvalue of $\mathbf{H}$ should converge to zero at a rate $\mathcal{O}[n^{-1/(d+2)}]$ (Liu, 2001).

Under assumptions (A1)–(A3), (H2), (K1) and (K2), the asymptotic conditional bias of estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 0)$, at a point $\mathbf{x}$ in the interior of the support of f, is:

$$\begin{aligned}\mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x};0) - m(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_m(\mathbf{x})] \\ &\quad + \frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}} m(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla} f(\mathbf{x}) \\ &\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)],\end{aligned}$$

and the asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{R(K)}{n|\mathbf{H}|f(\mathbf{x})}[\sigma^2 + f(\mathbf{x})\rho_c] + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \tag{1.30}$$

On the other hand, the asymptotic properties of the local linear estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, defined in (1.9), were also derived by Liu (2001) for short range dependence. Under assumptions (A1)–(A3), (H2), (K1) and (K2), the asymptotic conditional bias of the estimator $\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)$, at a point $\mathbf{x}$ in the interior of the support of f, is:

$$\mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x};1) - m(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_m(\mathbf{x})] + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)],$$

and the asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \;=\; \frac{R(K)}{n|\mathbf{H}|f(\mathbf{x})}[\sigma^2 + f(\mathbf{x})\rho_c] + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \quad (1.31)$$

As for independent data (see Section 1.1.1), it can be observed that the Nadaraya–Watson and local linear estimators of m for correlated data have the same asymptotic variance, but their asymptotic biases are different. As expected, the asymptotic bias for independent and for spatially correlated data are identical. Notice that for independent data, it follows that $\rho_c = 0$, and consequently the asymptotic variances of both Nadaraya–Watson and local linear estimators given in (1.30) and in (1.31), reduce to the asymptotic variance expressions for independent data given in (1.8) and (1.11), respectively.

Bandwidth selection

Following the same reasoning as for independent data to obtain (1.18) and (1.19), local optimal bandwidths could be also derived when the data are correlated. The optimal bandwidth matrix for the Nadaraya–Watson estimator (1.6) is:

$$\mathbf{H}^s_{\mathrm{opt}}(\mathbf{x};0) \;=\; \left\{\frac{R(K)[\sigma^2 + f(\mathbf{x})\rho_c]}{nd\mu_2^2(K)f(\mathbf{x})}|\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x})|^{1/2}\right\}^{1/d+4} \cdot \left[\tilde{\boldsymbol{\mathcal{G}}}(\mathbf{x})\right]^{-1/2}.$$

The optimal smoothing matrix to approximate the local linear estimator (1.9) is

$$\mathbf{H}^s_{\mathrm{opt}}(\mathbf{x};1) \;=\; \left\{\frac{R(K)[\sigma^2 + f(\mathbf{x})\rho_c]}{nd\mu_2^2(K)f(\mathbf{x})}|\tilde{\boldsymbol{\mathcal{H}}}_m(\mathbf{x})|^{1/2}\right\}^{1/d+4} \cdot \left[\tilde{\boldsymbol{\mathcal{H}}}_m(\mathbf{x})\right]^{-1/2}.$$

However, as it was stated for independent data, these optimal bandwidths depend also on unknown functions, such as the design density f, the Hessian matrix of the unknown regression function m and the error correlation. These quantities should be previously estimated for their use in practice. In order to obtain a global bandwidth, the optimization problem to solve is not trivial, not being possible to provide a closed form solution. Some simple cases were analyzed by Liu (2001). For univariate data, Francisco-Fernández and Vilar-Fernández (2001) proposed a plug-in global bandwidth selector, starting with a pilot bandwidth computed by using the time

series cross-validation criterion proposed by Hart (1994). For references on other methods for bandwidth selection in this context see Opsomer et al. (2001).

Alternatively, several bandwidth selection methods have been proposed in kernel regression estimation with dependent data. Most of the proposed procedures use a cross-validation algorithm. However, it should be noted that for correlated data, *simple* cross-validation criteria will not provide satisfactory bandwidths (Altman, 1990). For instance, if the errors are positively correlated, the leave-one-out cross-validation provides smaller bandwidths than the optimal one, leading to undersmoothing. In order to cope with possible dependence, several modifications of the cross-validation criterion has been proposed. For example, in the univariate case, Härdle and Vieu (1992) proposed the leave-$(2l+1)$-out version of cross-validation. The stronger is the dependence, larger l must be selected. Other smoothing parameter selection methods rely on estimating the correlation function and incorporating this estimate into the selection criterion (see Altman, 1990; Chiu, 1989; Hart, 1991). For spatially correlated data, Liu (2001) proposed a modified version of the cross-validation criterion (MCV). The smoothing parameter is chosen by selecting the matrix bandwidth $\mathbf{H}$ minimizing the function

$$\mathrm{MCV}(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z_i - \hat{m}_{\mathbf{H}}(\mathbf{X}_i;p)}{1-\mathbf{S}_{\mathbf{X}_i;p}\boldsymbol{\rho}_i}\right]^2, \tag{1.32}$$

where the vectors $\mathbf{S}_{\mathbf{X}_i;p}$, for $p=0,1$, were given in (1.23) and in (1.24), respectively, and $\boldsymbol{\rho}_i = [\rho_n(\mathbf{X}_1-\mathbf{X}_i),\ldots,\rho_n(\mathbf{X}_n-\mathbf{X}_i)]^{\mathrm{T}}$. For the modified version of GCV (MGCV), the matrix bandwidth is chosen by minimizing the function

$$\mathrm{MGCV}(\mathbf{H}) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z_i - \hat{m}_{\mathbf{H}}(\mathbf{X}_i;p)}{1-\frac{1}{n}\mathrm{tr}(\mathbf{S}_p\mathbf{R})}\right]^2, \tag{1.33}$$

where $\mathbf{R}$ is the $n\times n$ correlation matrix which ith row is $\boldsymbol{\rho}_i$.

Notice that the minimization of (1.32) and (1.33) requires knowledge of the correlation matrix of the errors, but not the variance. A MGCV bandwidth was proposed by Francisco-Fernandez and Opsomer (2005), considering a parametric specification of the correlation function. In particular, the authors used the isotropic exponential model given in (1.28), but other models could be employed instead.

1.2 Comparing parametric and nonparametric regression fits

Parametric and nonparametric regression estimation has been presented in Section 1.1 in different frameworks, for unidimensional and multidimensional covariate, considering independent and spatially correlated data.

As it was pointed out before, if the unknown regression function belongs to a known parametric family, then parametric procedures should be employed for regression estimation. However, parametric techniques may not provide satisfactory results if the model is not correctly specified. In order to assess the adequacy of a certain parametric form for the regression function, goodness-of-fit tests can be applied. The specific testing problem is formulated as:

$$H_0 : m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}, \qquad \text{vs.} \qquad H_a : m \notin \mathcal{M}_{\boldsymbol{\beta}}, \tag{1.34}$$

where $\boldsymbol{\mathcal{B}} \subset \mathbb{R}^q$ is a compact set, and q denotes the dimension of the parameter space $\boldsymbol{\mathcal{B}}$. For example, in the bidimensional case ($d = 2$), considering that $\mathcal{M}_{\boldsymbol{\beta}}$ is the family of linear models, then $q = 3$.

Test statistics are usually proposed by comparing a nonparametric pilot regression estimator and a corresponding parametric estimator of the regression function under the null hypothesis. For this purpose, tests based on empirical regression processes (Stute, 1997), on maximum likelihood ideas (see Fan et al., 2001), or on kernel-type methods for regression, have been designed and studied. For independent data, Härdle and Mammen (1993) proposed a test statistic to check if a regression function belongs to a class of parametric models by measuring the L_2-distance between a parametric and a nonparametric regression function estimates:

$$T_n = nh^{d/2} \int_{\mathcal{D}} [\hat{m}_h(\mathbf{x}) - \hat{m}_{h,\hat{\boldsymbol{\beta}}}(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x}, \tag{1.35}$$

where $\hat{m}_h(\mathbf{x})$ denotes a nonparametric estimator, $\hat{m}_{h,\hat{\boldsymbol{\beta}}}$ is a smoothed version of the parametric estimator $m_{\hat{\boldsymbol{\beta}}}$ and w is a weight function that helps in mitigating possible boundary effects. Specifically, the Nadaraya–Watson estimator was considered as a nonparametric approach in Härdle and Mammen (1993). A discretized version of

the test statistic given in (1.35) can be found in González-Manteiga and Cao (1993).

Similar tests to the one proposed by Härdle and Mammen (1993) have been studied considering other kernel-type regression estimators, with different estimators under the null hypothesis or using other discrepancy measures. For instance, Alcalá et al. (1999) proposed a testing procedure to check the parametric null hypothesis using local polynomial regression estimators. A supremum-norm-based test, comparing a kernel regression smoother with a parametric least squares estimator, was derived by Kozek (1991). In the context of time series, following the same strategy, goodness-of-fit tests for linear regression models with correlated errors have been proposed by González Manteiga and Vilar Fernández (1995) and Biedermann and Dette (2000). An exhaustive review on goodness-of-fit tests for regression models has been provided by González-Manteiga and Crujeiras (2013).

Although the problem of assessing a parametric regression model has been broadly studied for different frameworks, this is not the case for spatially correlated data. In this context, to decide if a parametric (like the one described in Section 1.1.2) or a nonparametric procedure must be chosen to estimate the unknown regression function, a goodness-of-fit test is studied in Chapter 2. Following similar ideas as those of Härdle and Mammen (1993), the proposed test statistic is based on a comparison between a kernel-type regression estimator and a $\sqrt{n}$-consistent estimator under the null hypothesis, using a weighted L_2-distance.

1.3 A linear-circular regression model

In the previous sections, some inference problems for regression models considering an independent or dependent $\mathbb{R}^d$-valued random covariate have been presented. However, in some occasions data may present certain *complexities*, involving some problems for the application of the previous estimation and testing procedures. Thus, it could happen that the response and/or the explanatory variables were of functional nature, or directional variables (in particular circular ones), or data with outliers, or interval-grouped data, or censored data, among other possible situations. In addition, within the context of this dissertation, it may happen that these *complex* observations exhibit spatial dependence. Among the possible regression models with

complex data, we will focus on multiple linear-circular regression models. More specifically, regression models with a circular response and an $\mathbb{R}^d$-valued predictor are considered:

$$\Theta = [m^c(\mathbf{X}) + \varepsilon](\mathtt{mod}\, 2\pi), \tag{1.36}$$

where Θ is a circular random variable taking values on $\mathbb{T} = [0, 2\pi)$, $\mathbf{X}$ is an Euclidean random variable supported on $\mathcal{D} \subseteq \mathbb{R}^d$, m^c is the circular regression function, ε is the circular random error, which is supposed to be zero mean direction, and `mod` stands for the modulo operation. The circular regression function m^c in model (1.36) is the conditional mean direction of Θ given $\mathbf{X}$ which, at a point $\mathbf{x}$, can be defined as the minimizer of the risk $\mathbb{E}\{1 - \cos[\Theta - m^c(\mathbf{X})] \mid \mathbf{X} = \mathbf{x}\}$.

A brief introduction to circular data is carried out to present notation and some special features of this kind of data. Circular descriptive statistics and some distributions on the circle are introduced.

Circular descriptive statistics

Circular data arise in many scientific fields where observations are recorded as directions or angles relative to a system with a fixed orientation. Once that a direction and a sense of rotation have been chosen, circular data can be expressed as angles (in degrees or in radians) or unit vectors on the circle. A complete introduction on circular data can be found in Mardia (1972) and Fisher (1995), or more recently, Mardia and Jupp (2000), Jammalamadaka and SenGupta (2001), or Ley and Verdebout (2017). Examples of circular data include wind directions (Fisher, 1995; Johnson and Wehrly, 1978), angles in the structure of a protein (Hamelryck et al., 2012), waves directions (Jona-Lasinio et al., 2012; Wang and Gelfand, 2014; Wang et al., 2015), animal orientations (Batschelet, 1981; Scapini et al., 2002; Schmidt-Koenig, 1963), arrival times (Cox and Lewis, 1996, pp. 254-255), and cyclical or seasonal patterns (Ameijeiras-Alonso et al., 2019), among others.

The circular nature of such data encompasses some challenges for applying traditional statistical methods used for Euclidean data. An example that directly illustrates this problem is the definition of sample mean. A first attempt could be to consider the classical Euclidean mean, however it is not appropriate. For example,

fixing a sense of rotation, the sample mean of $\pi/3$ and $5\pi/3$ would be equal to π, which provides an opposite direction of the mean, corresponding to 0. Given the circular sample $\{\Theta_1, \ldots, \Theta_n\}$ from a circular random variable Θ defined on $\mathbb{T} = [0, 2\pi)$, the circular sample mean is given by:

$$\bar{\Theta} = \text{atan2}\left[\frac{1}{n}\sum_{i=1}^{n}\sin(\Theta_i), \frac{1}{n}\sum_{i=1}^{n}\cos(\Theta_i)\right], \tag{1.37}$$

where the function $\text{atan2}(y, x)$ returns the angle between the x-axis and the vector from the origin to (x, y). The definition of a quadrant-specific inverse of the tangent is needed by the fact that $\tan(\theta) = \tan(\theta + \pi)$, so that there are two inverses for any given angle $\theta \in \mathbb{T}$. Since atan2 is defined to take values in $(-\pi/2, \pi/2)$, the above definition of the sample mean provides an unique inverse on $[0, 2\pi)$.

The sample mean resultant length, which is defined as:

$$\bar{R} = \frac{1}{n}\sqrt{\left[\sum_{i=1}^{n}\sin(\Theta_i)\right]^2 + \left[\sum_{i=1}^{n}\cos(\Theta_i)\right]^2},$$

is a useful measure of how concentrated the data are towards the mean direction, and takes values in $(0, 1)$. If all observations have the same direction, the variability is zero, and the resultant vector mean length is equal to one, which is the theoretical maximum. The sample circular variance can be defined as $V = 1 - \bar{R}$, and similarly to the variance of Euclidean data, the smaller the value of the sample circular variance, the more concentrated the data distribution. Notice that, unlike for Euclidean data, the circular variance can only takes values in $[0, 1]$ (Fisher, 1995).

In order to explore the relationship between circular variables, a correlation coefficient for circular data must be properly defined. Considering a pair of circular variables $\Theta, \Psi \in [0, 2\pi)$, and trying to retain many of the properties about the correlation coefficient for Euclidean data, Jammalamadaka and Sarma (1988) defined the circular correlation coefficient as follows:

$$r(\Theta, \Psi) = \frac{\mathbb{E}[\sin(\Theta - \mu_1)\sin(\Psi - \mu_2)]}{\sqrt{\mathbb{V}\text{ar}[\sin(\Theta - \mu_1)]\mathbb{V}\text{ar}[\sin(\Psi - \mu_2)]}}, \tag{1.38}$$

where μ_1 and μ_2 denote the mean directions of the circular variables Θ and Ψ, respectively. Notice that $\mathbb{E}[\sin(\Theta-\mu_1)] = \mathbb{E}[\sin(\Psi-\mu_2)] = 0$, and therefore, $\sin(\Theta-\mu_1)$ and $\sin(\Psi-\mu_2)$ can be taken to characterize the deviations of Θ and Ψ with respect their mean directions μ_1 and μ_2. The circular correlation coefficient r given in (1.38) satisfies the following properties:

- $r(\Theta,\Psi) = r(\Psi,\Theta)$, $\forall\Theta,\Psi \in [0,2\pi)$
- $|r(\Theta,\Psi)| \leq 1$, $\forall\Theta \in [0,2\pi)$
- $r(\Theta,\Psi) = 0$ if Θ and Ψ are independent, $\forall\Theta,\Psi \in [0,2\pi)$
- $r(\Theta,\Psi) = 1$ if and only if $\Theta = \Psi + k_1(\mathtt{mod}\, 2\pi)$, and $r(\Theta,\Psi) = -1$ if and only if $\Theta + \Psi = k_2(\mathtt{mod}\, 2\pi)$, $\forall\Theta,\Psi \in [0,2\pi)$, $k_1, k_2 \in \mathbb{R}$

As for Euclidean data, the sample circular correlation coefficient can be also defined. Let $\{(\Theta_i,\Psi_i)\}_{i=1}^n$ be a random sample of (Θ,Ψ), where Θ and Ψ are circular random variables. The sample correlation coefficient is given by:

$$r_n(\Theta,\Psi) = \frac{\sum_{i=1}^n \sin(\Theta_i-\bar{\Theta})\sin(\Psi_i-\bar{\Psi})}{\sqrt{\sum_{i=1}^n \sin^2(\Theta_i-\bar{\Theta})\sin^2(\Psi_i-\bar{\Psi})}}, \tag{1.39}$$

where $\bar{\Theta}$ and $\bar{\Psi}$ are the sample mean directions of Θ and Ψ in (1.37), respectively. The sample correlation coefficient r_n is an estimator of the circular correlation coefficient r (Jammalamadaka and Sarma, 1988). For further details on descriptive circular statistics we refer to Section 1.3. of Jammalamadaka and SenGupta (2001).

Some distributions on the circle

Circular data can be differently distributed on the circle. For instance, data can be uniform (all directions are equally likely), unimodal (when there is a single cluster of data points) or multimodal (when there exist two or more clusters in the data). A common unimodal distribution is the von Mises distribution, which can be seen as the Gaussian analogue for circular data, because it is the only circular distribution whose maximum likelihood estimator of the location parameter is the circular sample mean (Bingham and Mardia, 1975), as it occurs for the Gaussian distribution in

the Euclidean setting. The von Mises distribution and other well-known circular distributions will be introduced below.

From now on, to facilitate the explanation for circular variables, an abuse of notation will be made, denoting by f and μ the circular density function and the circular mean direction. Taking into account the periodicity, the circular density function f must satisfy the following assumptions (see Jammalamadaka and SenGupta, 2001, Chapter 2):

- $f(\theta) \geq 0$, $\forall \theta \in [0, 2\pi)$
- $\int_0^{2\pi} f(\theta) d\theta = 1$
- $f(\theta) = f(\theta + 2k\pi)$, $\forall \theta \in [0, 2\pi)$, $\forall k \in \mathbb{Z}$

Most of the classical distributions are characterized by the mean and concentration parameters. Large values of the concentration parameter indicate that the distribution is more concentrated around the mean parameter.

The von Mises distribution, $vM(\mu, \kappa)$, is a symmetric unimodal distribution with density function given by:

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)}, \tag{1.40}$$

where $\mu \in [0, 2\pi)$ is the mean direction, $\kappa \geq 0$ is the concentration parameter and I_0 is the modified Bessel function of the first kind and order zero, defined as $I_0 = (1/2\pi) \int_0^{2\pi} \exp[\kappa \cos(\theta)] d\theta$. Figure 1.4 (left) shows the von Mises density function with $\mu = \pi$ and $\kappa = 0.1$ (solid line), $\kappa = 2$ (dashed line) and $\kappa = 10$ (dotted line). Notice that, if κ is equal to zero, a particular case of the von Mises family is the circular uniform distribution.

Another way to obtain circular distributions is by wrapping a linear distribution around the circumference of unit radius. That is to say, if X is a real-valued random variable, the corresponding circular random variable can be obtained as $\Theta = X(\mathtt{mod}\, 2\pi)$. If g is the density of X, then the density of Θ is obtained as:

$$f(\theta) = \sum_{k=-\infty}^{\infty} g(\theta + 2\pi k), \quad \theta \in [0, 2\pi).$$

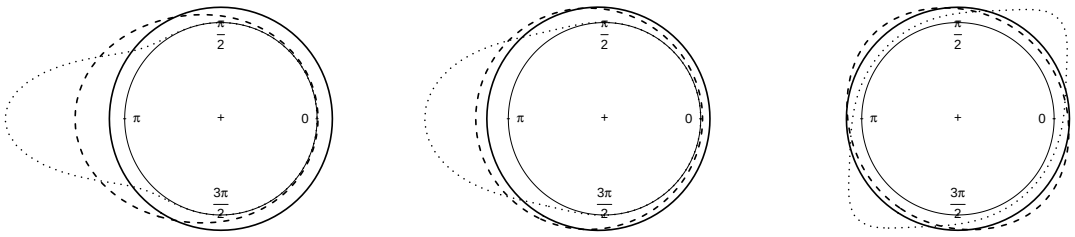

Figure 1.4: Examples of circular probability density functions. From left to right, representation of the models with different parameters: von Mises, with $\mu = \pi$ and $\kappa = 0.1$ (solid line), $\kappa = 2$ (dashed line) and $\kappa = 10$ (dotted line); wrapped normal, with $\mu = \pi$ and $\rho = 0.2$ (solid line), $\rho = 0.5$ (dashed line) and $\rho = 0.9$ (dotted line); and projected normal, with $\boldsymbol{\mu} = (0, 0)$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0$ (solid line), $\rho = -0.4$ (dashed line) and $\rho = 0.8$ (dotted line).

Some useful distributions on the circle can be obtained using this method (Mardia, 1972). For example, the wrapped Cauchy distribution, $WC(\mu, \rho)$, is a symmetric unimodal distribution which can be obtained by wrapping the Cauchy distribution, with location parameter μ and scale parameter $-\log(\rho)$, around the circle, where log stands for the natural logarithm. The probability density function has the following expression:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos(\theta - \mu)}.$$

The wrapped normal distribution, $WN(\mu, \rho)$, is constructed by wrapping the normal distribution, centered in μ with variance $-2\log(\rho)$, onto the circle. Its probability density function is:

$$f(\theta; \mu, \rho) = \frac{1}{2\pi} \left[1 + 2 \sum_{k=1}^{\infty} \rho^{k^2} \cos k(\theta - \mu) \right].$$

Figure 1.4 (center) shows the wrapped normal density function with $\mu = \pi$ and $\rho = 0.2$ (solid line), $\rho = 0.5$ (dashed line) and $\rho = 0.9$ (dotted line). For both wrapped Cauchy and normal distributions, if ρ is equal to zero, this leads to the circular uniform distribution.

Circular distributions can be also generated by projecting a bivariate distribution on a unit circle (Mardia and Jupp, 2000). For instance, the projected normal distribution is obtained by a radial projection of the distribution of a bivariate normal random variable $\mathbf{X} \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Xi})$, with $\boldsymbol{\mu} = (\mu_1, \mu_2)^{\mathrm{T}}$ and

$$\boldsymbol{\Xi} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \tag{1.41}$$

on the circle (Mardia, 1972). Taking $\mathbf{U} = (U_1, U_2) = \frac{\mathbf{X}}{\|\mathbf{X}\|}$, a circular random variable can be obtained defining Θ such that $U_1 = \cos\Theta$ and $U_2 = \sin\Theta$. Its probability density function is:

$$f(\theta) = \frac{\phi_2(\mu_1, \mu_2; 0, \boldsymbol{\Xi})}{C(\theta)} + \frac{aD(\theta)\Phi_1[D(\theta)]\phi_1[aC^{-1/2}(\theta)E(\theta)]}{C(\theta)},$$

where $a = (\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}$, $C(\theta) = a^2\left[\sigma_2^2\cos^2\theta - \rho\sigma_1\sigma_2\sin(2\theta) + \sigma_1^2\sin^2\theta\right]$, $D(\theta) = a^2C^{-1/2}(\theta)[\mu_1\sigma_2(\sigma_2\cos\theta - \rho\sigma_1\sin\theta) + \mu_2\sigma_1(\sigma_1\sin\theta - \rho\sigma_2\cos\theta)]$, $E(\theta) = \mu_1\sin\theta - \mu_2\cos\theta$ and, ϕ_1 and Φ_1 are the density and the distribution of a normal variable. This distribution was discussed by Mardia and Jupp (2000) and a good review was provided by Wang and Gelfand (2014). Notice that although the density function has a complicated expression, some specific features can be controlled for certain parameter values. For instance, if $\boldsymbol{\Xi} = \mathbf{I}_2$, the corresponding density function is symmetric and unimodal. Moreover, if $\boldsymbol{\Xi} = \mathbf{I}_2$ and $\mu_1 = \mu_2 = 0$, the distribution is uniform on the circle. Finally, if $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = \sigma$, the density is reduced to

$$f(\theta) = \frac{\sqrt{1-\rho^2}}{2\pi[1 - \rho\sin(2\theta)]}.$$

Figure 1.4 (right) shows the projected normal density function with $\boldsymbol{\mu} = (0, 0)$, $\sigma_1 = \sigma_2 = 1$ and $\rho = 0$ (solid line), $\rho = -0.4$ (dashed line) and $\rho = 0.8$ (dotted line).

Taking into account that circular data require a different treatment to that given to Euclidean data, it is necessary to properly adapt the existing inference procedures or to design new ones accounting for the particular nature of the data. In Chapter 3, 4 and 5 of this dissertation, inference problems for multiple linear-circular regression models (circular response and Euclidean covariates) will be discussed. More

specifically, in Chapters 3 and 4, nonparametric estimators of the circular regression function m^c in model (1.36), for independent and spatially correlated data, respectively, are proposed and studied. Chapter 5 is devoted to present some proposals on goodness-of-fit tests for parametric circular regression models for both independent and spatially correlated data.

1.4 Manuscript organization

In this section a brief summary of each chapter of the thesis will be provided. The aims are presented, as well as the chapter distribution. The contributions of this thesis are also highlighted in each chapter.

In Chapter 2, the problem of assessing a parametric regression model in the presence of spatial correlation is addressed. For that purpose, a goodness-of-fit test based on a L_2-distance comparing a parametric and a nonparametric regression estimator is proposed. Asymptotic properties of the test statistic, both under the null hypothesis and under local alternatives, are derived. Additionally, three bootstrap procedures are designed to calibrate the test in practice. Finite sample performance of the test is analyzed through an extensive simulation study, comparing the proposed bootstrap procedures and accounting for different features that usually appear in spatial analysis. An illustration with a real dataset is also provided. The contributions of this chapter have been collected in Meilán-Vila et al. (2020e) and Meilán-Vila et al. (2020b). In Meilán-Vila et al. (2020e), a random design and the multivariate local linear regression estimator for the nonparametric alternative, were considered. On the other hand, in Meilán-Vila et al. (2020b), under fixed design and taking the Nadaraya–Watson estimator as a nonparametric fit, a detailed computational analysis of the behavior of this test when proposing different bootstrap algorithms was provided.

Chapter 3 is devoted to the introduction and analysis of nonparametric estimators of a regression function in a model with a circular response and an $\mathbb{R}^d$-valued random predictor. Expressions for their asymptotic biases and variances are derived, and some guidelines to select asymptotically local optimal bandwidth matrices are also given. The finite sample behavior of the proposed estimators is assessed through

simulations and their performance is also illustrated with a real dataset. The contributions of this chapter can be found in Meilán-Vila et al. (2020d). When the response variable is circular, the regression function is given by the inverse tangent function of the ratio between the conditional expectation of the sine and the conditional expectation of the cosine of the response variable. The proposal considers two (separate) regression models for the sine and cosine components, which are indeed regression models with real-valued response. Then, nonparametric estimators for the circular regression function are obtained by computing the inverse tangent function of the ratio of multivariate local polynomial estimators for the sine and cosine models.

In Chapter 4, a regression model with a circular response and an $\mathbb{R}^d$-valued random predictor, assuming that the errors exhibit spatial correlation, is considered. Using a nonparametric approach, local polynomial type estimators for the circular regression function are proposed and studied. Their asymptotic bias and variance are derived. Finite sample performance of the estimators is analyzed through a simulation study, using wrapped and projected approaches to generate the spatially-dependent circular errors. A real data illustration is also presented. The contributions in this topic have been collected in Meilán-Vila et al. (2020a).

Chapter 5 is devoted to formulate goodness-of-fit tests for circular regression models. Test statistics are proposed to check if the circular regression function belongs to a known parametric family, comparing a (non-smoothed or smoothed) parametric fit with a nonparametric estimator of the circular regression function, using a circular distance. Appropriate bootstrap algorithms are designed to calibrate the tests in practice, both for independent and for spatially correlated data. The finite sample behavior of the procedures is checked through a simulation study. The tests are also applied to real datasets. The contributions of this chapter can be found in Meilán-Vila et al. (2020c).

The manuscript contains some comments and discussion in Chapter 6. Finally, an appendix collecting some useful theoretical results is also included.

Chapter 2

Testing parametric regression models with spatially correlated errors

2.1 Introduction

The problem of testing a parametric regression model, confronting a parametric estimator of the regression function with a smooth alternative estimated by a nonparametric method, has been approached by several authors in the statistical literature (see, for example Azzalini et al., 1989; Eubank and Spiegelman, 1990). For instance, Weihrather (1993) and Eubank et al. (2005) described tests based on an overall distance between parametric and nonparametric regression fits, giving some strategies on bandwidth selection. Härdle and Mammen (1993) proposed a testing procedure to check if a regression function belongs to a class of parametric models by measuring a L_2-distance between parametric and nonparametric estimates. Specifically, the Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964) was considered for the nonparametric approach. The same type of study was performed by Alcalá et al. (1999), but using a local polynomial regression estimator (Fan and Gijbels, 1996). Following similar ideas, a local test for a univariate parametric model checking was proposed by Opsomer and Francisco-Fernández (2010), while Li (2005) assessed the lack of fit of a nonlinear regression model, comparing a local linear smoother and

parametric fits.

The previous testing procedures, all of them formulated with independent errors, have been also adapted for scenarios where data exhibit correlation in time. For example, Park et al. (2015) considered a model specification test based on a kernel for a nonparametric regression model with an equally-spaced fixed design and correlated errors. Also in the context of time series, goodness-of-fit tests for linear regression models with correlated errors have been studied by González Manteiga and Vilar Fernández (1995), also considering an equispaced fixed design. Biedermann and Dette (2000) extended the previous results under fixed alternatives, considering a regression model with explanatory variables x_i, $i = 1, \ldots, n$, being fixed and given by $i/n = \int_0^{x_i} f(t)dt$, where f is a positive density on the interval $[0, 1]$. For further discussion and examples of nonparametric specification tests for regression models, see the comprehensive review by González-Manteiga and Crujeiras (2013).

Although for time dependent errors, the problem of assessing a parametric regression model has been widely studied, this is not the case for spatially correlated data (or even with spatio-temporal correlation). Observations from spatially varying processes are quite frequent in applied sciences such as ecology, environmental and soil sciences. In order to gain some insight in the process evolution across space, a regression model where the regression function captures the first-order structure, whereas the error term collects the second-order structure, can be formulated in the previous contexts. Usually, parametric models are considered for the regression function, e.g. polynomial models on latitude and longitude (see Cressie, 1993; Diggle and Ribeiro, 2007), and estimation is accomplished by least squares methods, providing reliable inferences if the model is correctly specified. As an example, a classical dataset which is analyzed under this scope is the Wolfcamp aquifer data presented by Harper and Furr (1986), collecting 85 measurements of levels of piezometric-head (see Figure 2.1). In this example, several parametric trend models are considered after performing different analyses, concluding that a linear trend seems to be a reasonable model (see Figure 2.2). However, to determine if this linear model (or in general, any parametric fit) is an appropriate representation of a dataset, it would be advisable to carry out a statistical test in order to assess the goodness-of-fit of the selected model. In this context, the statistical literature initially focused on the

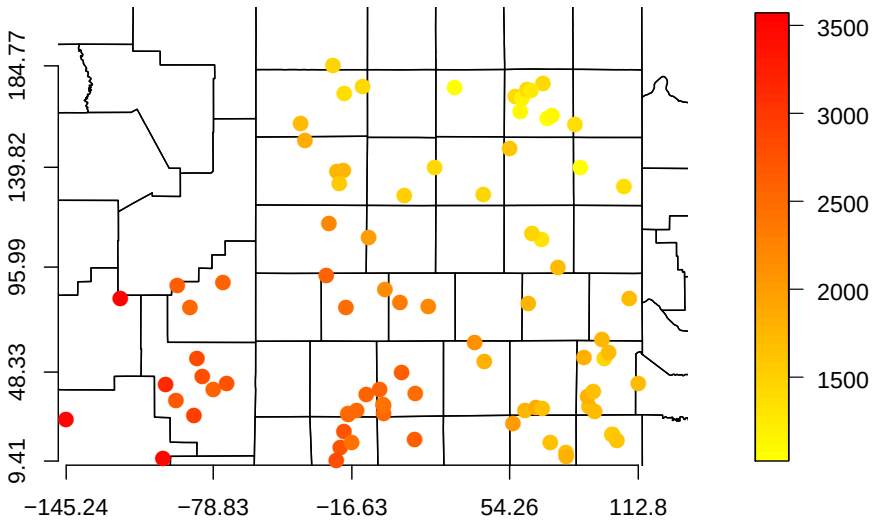

Figure 2.1: Locations with the levels of piezometric-head for the Wolfcamp Aquifer.

assessment of independence (Diblasi and Bowman, 2001) and on testing a parametric correlation model (Maglione and Diblasi, 2004), considering the variogram as the function describing the spatial dependence pattern. Also taking the variogram as the target function, Bowman and Crujeiras (2013) proposed some testing methods for simplifying hypothesis (namely, stationarity and isotropy). Although these proposals investigate the dependence structure of the data (a nuisance when the primary goal is the regression or trend function), the ideas which inspired these methods are common to the goodness-of-fit tests for regression models.

A new proposal for testing a parametric regression model (with univariate responses and possibly d-dimensional covariates), for spatially correlated data, is presented in this chapter. Following similar ideas as those of Härdle and Mammen (1993), the test statistic is based on a comparison between a smoothed version of a parametric fit and a nonparametric estimator of the regression function, using a weighted L_2-distance. The null hypothesis that the regression function follows a parametric model is rejected if the distance exceeds a certain threshold. To perform the parametric estimation, an iterative procedure based on generalized least squares is used (see Diggle and Ribeiro, 2007), although other fitting techniques such as maximum likelihood methods could be employed. For the nonparametric alternative, the multivariate Nadaraya–Watson or local linear regression estimator is

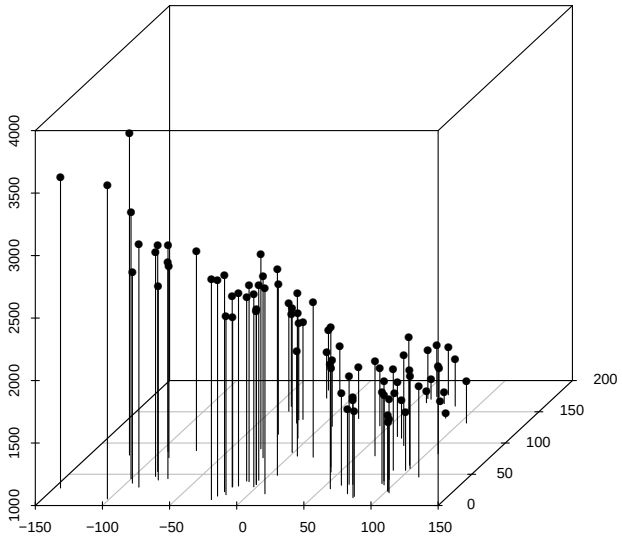

Figure 2.2: 3-dimensional representation of the levels of piezometric-head for the Wolfcamp Aquifer.

used (Francisco-Fernandez and Opsomer, 2005; Hallin et al., 2004; Härdle and Müller, 2012; Liu, 2001), generalizing in some way the results of Härdle and Mammen (1993) and Alcalá et al. (1999) for the univariate case with independent errors.

The proposed test statistic shows a slow rate of convergence to its asymptotic distribution, motivating the use of resampling methods to approximate its distribution under the parametric null hypothesis. It should be noted that, in order to mimic the process behavior under the null hypothesis, not only the parametric form of the regression function has to be considered, but also the (unspecified) spatial dependence of the data, which has to be recovered from a single realization of the spatial process, under stationarity conditions. In the presence of spatial correlation, resampling methods may not be accurate enough for mimicking the spatial dependence structure under the null hypothesis from a single realization of the process. This is the reason why a thorough analysis of the impact of the spatial dependence configuration in the distribution approximation is required and provided.

Traditional resampling procedures for test calibration designed for independent data should not be used for spatial processes, as they do not account for the cor-

relation structure. One of the aims of this chapter is to present and analyze three different proposals for test calibration which take the dependence of the data into account: a parametric residual bootstrap (PB), a nonparametric residual bootstrap (NPB) and a bias-corrected nonparametric bootstrap (CNPB). Parametric bootstrap procedures, following the ideas introduced by Solow (1985), are a usual strategy in spatial statistics, since they directly involve the dependence structure (see, for example, Olea and Pardo-Iguzquiza, 2011). The PB approach consists in using in the bootstrap algorithm the residuals obtained from the parametric fit and, from these residuals, estimating parametrically the spatial dependence structure. If the regression function indeed belongs to the parametric family considered in the null hypothesis, then the residuals obtained with this approach will be *similar* to the theoretical errors, and it is expected that the PB method will have a good performance. A possible drawback of this procedure is the misspecification of the parametric model selected for the dependence estimation, however this issue could be avoided by using a nonparametric estimator instead. Moreover, this resampling approach relies on the wrong assumption that the variability of the residuals is the same as the one of the theoretical errors. In the NPB method, to increase the power of the test, residuals are obtained from the nonparametric fit (see González-Manteiga and Cao, 1993). Furthermore, the dependence structure is estimated without considering parametric assumptions. It is clear that the NPB resamplig method can avoid the misspecification problem both for the regression function and the dependence. However, no matter the method used to remove the first-order structure, either parametric or nonparametric, the direct use of residuals gives rise to biased variogram estimates, especially at large lags (see Cressie, 1993, Section 3.4.3). To solve this problem, the CNPB approach is a modification of the NPB method, but including a bias-corrected algorithm for the dependence estimation (see Castillo-Páez et al., 2019; Fernández-Casal and Francisco-Fernández, 2014).

This chapter is organized as follows. Section 2.2 introduces the testing problem, as well as the proposed L_2-test statistic. Required assumptions and the asymptotic distribution of the test statistic are also presented. A detailed description of the calibration algorithms considered is given in Section 2.3. In addition, an exhaustive simulation study for assessing the performance of the test with PB, NPB and CNPB,

is presented in Section 2.4. Section 2.5 shows how to apply the testing procedure to the Wolfcamp aquifer dataset. Finally, Section 2.6 includes the proofs of the main results.

2.2 A goodness-of-fit test for parametric regression models with spatially correlated errors

In model (1.1), the regression function m can be characterized using parametric or nonparametric models. As pointed out in Chapter 1, parametric models are usually easy to compute and provide for a direct interpretation of the parameter values. On the other hand, nonparametric models also provide a global view of the mean of the process. Their flexibility allows to model complex relations beyond a parametric form. Therefore, a question of interest in spatial modeling is focused on characterizing the first order structure of the process, checking if the regression function belongs to a parametric family by solving the testing problem given in (1.34), where $m_{\boldsymbol{\beta}}$ denotes a d-variate parametric function with parameter vector $\boldsymbol{\beta}$. Note that $m_{\boldsymbol{\beta}}$ is not restricted to be polynomial, although that is a common choice in practice.

To tackle this problem, a natural approach consists in comparing a parametric estimator of the regression function with a nonparametric one. The question arises if the differences between both fits can be explained by small stochastic fluctuations or if such differences suggest that the parametric assumption is not correct and it is more reasonable to use nonparametric methods to approximate the regression function. Using these ideas, one way to proceed is to measure the distance between both fits and to employ this distance as the test statistic for checking the parametric model. The estimation methods (parametric and nonparametric) considered in this proposal has been described in Section 1.1.2. Notice that the parametric estimator used in the test must satisfy a $\sqrt{n}$-consistency property. Moreover, as pointed out in Section 1.1.2, a note of caution should be made about regression estimation in this context: for spatially correlated data, when just a single realization of the process is

available, additional stationarity assumptions on the process are required in order to enable statistical inference. In addition, from a single realization, it may be difficult to disentangle the regression and error components, especially if the dependence is strong.

The approach followed to solve the testing problem (1.34), as in Härdle and Mammen (1993) or Alcalá et al. (1999), considers a test statistic given by a weighted L_2-distance between the nonparametric and parametric fits:

$$T_{n,p} = n|\mathbf{H}|^{1/2} \int_{\mathcal{D}} [\hat{m}_{\mathbf{H}}(\mathbf{x}; p) - \hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)]^2 w(\mathbf{x}) d\mathbf{x}, \tag{2.1}$$

for $p = 0, 1$, where w is a weight function that helps in mitigating possible boundary effects. The use of a weight function is quite frequent in this type of tests, both for density and regression (González-Manteiga and Crujeiras, 2013). The estimators $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$, for $p = 0, 1$, are the Nadaraya–Watson or local linear fits of m given in (1.6) or in (1.9), respectively. Moreover, $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 0)$ and $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 1)$ are smoothed versions of the parametric estimator $m_{\hat{\boldsymbol{\beta}}}$ constructed using (1.27). The smoothers are defined as:

$$\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 0) = \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})},$$

and

$$\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 1) = \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \mathbf{m}_{\hat{\boldsymbol{\beta}}}, \tag{2.2}$$

with $\mathbf{m}_{\hat{\boldsymbol{\beta}}} = [m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_1), \ldots, m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_n)]^{\mathrm{T}}$.

In the particular situation that the parametric family $\mathcal{M}_{\boldsymbol{\beta}}$ in (1.34) is the class of polynomials of degree less or equal than k, it could be more reasonable to use, as the nonparametric fit, the multivariate local polynomial estimator of degree l, with $l \geq k$, and considering the L_2-distance between this estimator and $m_{\hat{\boldsymbol{\beta}}}$. In that case, it would not be necessary to employ a smoothed version of $m_{\hat{\boldsymbol{\beta}}}$, because both are consistent unbiased estimators of the regression function, under the null hypothesis. However, for a general parametric family $\mathcal{M}_{\boldsymbol{\beta}}$, this is not true. For instance, using the local linear estimator, given that $\mathbb{E}[\hat{m}_{\mathbf{H}}(\mathbf{x}; 1)] = \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} m(\mathbf{x})$, it is convenient to smooth the parametric estimator so that the parametric term in (2.1) has the same expected value as the nonparametric term, under H_0. This fact

also justifies the use of the same bandwidth matrix $\mathbf{H}$ in the nonparametric estimator and in the smoothed version of the parametric fit (see Härdle and Mammen, 1993, page 1928).

If the null hypothesis is true, then the parametric and nonparametric estimators in (2.1) will tend to be *similar* and the value of $T_{n,p}$ will be *small*. Conversely, if the null hypothesis is does not hold, major differences between both fits will be expected, and therefore, the value of $T_{n,p}$ will be large. So, H_0 will be rejected if the distance between both fits exceeds a critical value. For example, consider the Wolfcamp aquifer dataset described in Section 2.1. Figure 2.8 (in Section 2.5) shows the smoothed version of the parametric (left) and the nonparametric (right) regression estimators for the level of piezometric-head in the area of study. In this case, a linear model is considered for the parametric fit, while the local linear estimator (1.9) is employed to perform the nonparametric fit. A multiplicative triweight kernel and the optimal bandwidth obtained by minimizing the mean average squared error (MASE) of the local linear estimator (see Francisco-Fernandez and Opsomer, 2005, page 288) are considered to compute $\hat{m}_{\mathbf{H}}(\mathbf{x};1)$ and $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};1)$. In this case, from a visual comparison, one may argue that given that both estimates are very similar, the value of the test statistic $T_{n,1}$ is *small*, and consequently, there may be no evidences against the assumption of a linear trend. However, apart from getting some insight to what might occur when using exploratory methods, in order to formally test the model using $T_{n,p}$ given in (2.1), it is essential to approximate the distribution of the test statistic under the null hypothesis.

The types of model deviations that are captured by this test are of the form $m(\mathbf{x}) = m_{\boldsymbol{\beta}_0}(\mathbf{x}) + c_n g(\mathbf{x})$, where c_n is a sequence, such that $c_n \to 0$ and g is a deterministic function collecting the deviation direction from the null model. In this section, the asymptotic distribution of the test statistic (2.1) is derived under the null hypothesis, and also under local alternatives converging to the null hypothesis at a certain rate controlled by c_n. Specifically, it is assumed that the function g is bounded (uniformly in $\mathbf{x}$ and n) and $c_n = n^{-1/2}|\mathbf{H}|^{-1/4}$. In particular, this contains the null hypothesis corresponding to $g(\mathbf{x}) = 0$.

It is clear from expression (2.1) that $T_{n,p}$ depends on the bandwidth matrix $\mathbf{H}$. A non-trivial problem in goodness-of-fit testing is the bandwidth choice, since the

optimal bandwidth for estimation may not be the optimal one for testing (being not even clear what optimal means). For instance, Fan et al. (2001), Eubank et al. (2005) and Hart (2013) gave some strategies on bandwidth selection in testing problems. This issue was also discussed further in detail by Sperlich (2013). As usual in the context of smooth-based goodness-of-fit tests for regression models, the performance of the test statistic $T_{n,p}$ is analyzed for a range of bandwidths in the numerical studies, allowing to check how sensitive the results are to variations in $\mathbf{H}$. Note that although technically it is possible to consider different bandwidth matrices in $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)$, the use of just one bandwidth matrix simplifies the application of the test in practice.

Note that the test statistic (2.1) generalizes to the framework of spatial correlated data (with a d-dimensional covariate) the statistic proposed for independent data by Härdle and Mammen (1993), using the Nadaraya–Watson estimator, and that of Alcalá et al. (1999) using the local polynomial estimator and considering a single covariate.

Next, the asymptotic distribution of $T_{n,p}$ is derived. Apart from some of the conditions stated in Chapter 1 for nonparametric kernel estimators (that will be specified below), the following assumptions on the stochastic nature of the observations are needed:

(A4) The covariate $\mathbf{X}$ lies in a compact set with probability one. The marginal density f is bounded away from zero.

(A5) The weight function w is continuously differentiable.

(A6) For any i, j, k, l,

$$\mathbb{C}\mathrm{ov}(\varepsilon_i\varepsilon_j, \varepsilon_k\varepsilon_l) = \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_k)\mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_l) + \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_l)\mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_k).$$

(A7) It is assumed that errors are a geometrically strong mixing sequence with mean zero and $\mathbb{E}|\varepsilon(\mathbf{x})|^r < \infty$ for all $r > 4$.

Assumption (A6) is satisfied, for example, when the errors follow a Gaussian distribution. As for (A7), if $\mathcal{M}_a^b$ is the σ-field generated by $\{\xi(t) : a \leq t \leq b\}$, then

$\{\xi(t) : t \in \mathbb{R}\}$ is geometrically strong mixing if the mixing coefficients verify:

$$\alpha(\tau) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{M}_{-\infty}^{0} \quad \text{and} \quad B \in \mathcal{M}_{\tau}^{\infty}\} = \mathcal{O}(\zeta^{\tau}), \quad (2.3)$$

for some $0 < \zeta < 1$, when $\tau \to \infty$. This assumption is needed to apply the central limit theorem for reduced U-statistics under dependence given by Kim et al. (2013). Note that if a random variable is a real Gaussian process, the strong mixing coefficient and the correlation function are equivalent (Rozanov, 1967, page 181). Therefore, hypotheses (A3), (A6) and (A7) could be satisfied by Gaussian error processes with exponential or rational quadratic (among others) correlation functions, given in (1.28) and in (1.29), respectively, having a decay rate larger than or equal to that indicated in (2.3).

As pointed out before, for the parametric estimator, just the assumption of being a $\sqrt{n}$-consistent estimator is required. This is guaranteed if the parametric estimator $m_{\hat{\boldsymbol{\beta}}}$ described in Section 1.1.2 is employed in the statistic (2.1). Anyway, a different parametric estimator of the regression function could be used in the test statistic (2.1) as long as this property was fulfilled.

The following theorem shows the asymptotic distribution of the test statistic proposed in (2.1).

Theorem 2.1 *Under assumptions* (A1)–(A7), (H2), (K1) and (K2)*, and if* $0 < V < \infty$*, it can be proved that for* $p = 0, 1$*,*

$$V^{-1/2}(T_{n,p} - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

where $\to_{\mathcal{L}}$ *denotes convergence in distribution, with*

$$\begin{aligned} b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2}\sigma^2 K^{[2]}(\mathbf{0})\left[\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right], \\ b_{1\mathbf{H}} &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x}, \end{aligned}$$

and

$$V = 2\sigma^4 K^{[4]}(\mathbf{0})\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right],$$

where $K^{[j]}$ *denotes the* j*-times convolution product of* K *with itself.*

Proof See Section 2.6.

This result generalizes to spatial correlated data the asymptotic distribution obtained for independent data (corresponding to $\rho_c = 0$) by Härdle and Mammen (1993), using the Nadaraya–Watson estimator, and that of Alcalá et al. (1999), considering the local polynomial regression estimator and a single covariate. From Theorem 2.1 it can be obtained that the asymptotic distribution of the test is the same if Nadaraya–Watson or local linear is employed in the nonparametric fit.

Geostatistical spatial trend models

The asymptotic distribution of the test statistic (2.1) can be also obtained under a geostatistical spatial trend model.

In this scenario, model (1.1) can be viewed as an additive decomposition of the spatial process: the regression or trend function m corresponds to the first-order moment of the process and captures the large-scale variability, whereas the error term collects the second-order structure, reflecting the small-scale variation. The covariates in this setting are given by the spatial locations (latitude and longitude), which are usually fixed in a geostatistical setting. In this case, considering assumptions in Theorem 2.1, except the ones relative to f (given that we are under a fixed design scheme), and following similar steps to those employed in the proof of Theorem 2.1, but using Riemann approximations of sums by integrals, the asymptotic distribution of $T_{n,p}$ is given in the following result:

Theorem 2.2 *Under assumptions* (A2), (A5)–(A7), (H2), (K1) and (K2)*, and if* $0 < V < \infty$*, it can be proved that for* $p = 0, 1$*,*

$$V^{-1/2}(T_{n,p} - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

with

$$\begin{aligned} b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2}\sigma^2 K^{[2]}(\mathbf{0})\left[\int w(\mathbf{x})d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right], \\ b_{1\mathbf{H}} &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x}, \end{aligned}$$

and

$$V = 2\sigma^4 K^{[4]}(\mathbf{0})\left[\int w^2(\mathbf{x})d\mathbf{x} + 2\rho_c \int w^2(\mathbf{x})d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right].$$

Proof See Section 2.6.

For the sake of simplicity, in model (1.1), it is considered that the nugget is equal to zero, and Theorems 2.1 and 2.2 focus on this case. Models with nugget different from zero are analyzed through simulations.

2.3 Test statistic calibration

The asymptotic distribution of the test obtained in Theorems 2.1 and 2.2, as in other nonparametric testing procedures (see, for example, Härdle and Mammen, 1993), may not be sufficiently precise to approximate the test statistic distribution under the null hypothesis in practice, for small or moderate sample sizes. Given the slow rate of convergence, to obtain an accurate approximation of the asymptotic distribution of the test, it would be necessary to have a very large sample size, which is not always the case for spatial data. Moreover, the limit distribution of the test statistic depends on unknown quantities that must be estimated. This is a common problem in smoothing-based tests, as already noted by González-Manteiga and Crujeiras (2013). This issue is usually overcome using resampling methods, specifically, employing bootstrap algorithms that try to mimic the data structure under the null hypothesis. Nevertheless, and for the sake of illustration, a brief simulation experiment is presented to study the performance of the asymptotic distribution of the test under the null hypothesis in practice.

We consider the simple case of assuming f and σ^2 known, and the density estimator of $V^{-1/2}(T_{n,1} - b_{0\mathbf{H}})$ and the standard normal density function are compared. The parametric regression family:

$$\mathcal{M}_{1,\beta} = \{\beta_0 + \beta_1 X_1 + \beta_2 X_2, \beta_0, \beta_1, \beta_2 \in \mathbb{R}\} \tag{2.4}$$

is assumed for the null hypothesis. In this case, the chosen regression function has the expression:

$$m(\mathbf{X}) = X_1 + X_2, \quad \mathbf{X} = (X_1, X_2). \tag{2.5}$$

In this experiment, 500 samples of size $n = 400$, 2500 and 10000 are generated from a regression model with explanatory variables drawn from a bivariate uniform distribution in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$ and regression function (2.5). The random errors ε_i are normally distributed with zero mean and with isotropic exponential covariance function:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \sigma^2[\exp(-\lambda n \|\mathbf{X}_i - \mathbf{X}_j\|)], \tag{2.6}$$

with values of $\sigma^2 = 1$ and $\lambda = 0.0005$ (for the sake of simplicity, no nugget effect is considered in this experiment). Note that with this selection of λ, the values for the practical range are 5, 0.8 and 0.2, for $n = 400$, 2500 and 10000, respectively. The parametric fit was computed using the iterative least squares procedure described in Section 1.1.2, considering a linear model. The nonparametric fit was obtained using the multivariate local linear estimator with a multivariate Gaussian kernel and a scalar bandwidth matrix. With this kernel, the quantities $K^{[2]}(\mathbf{0})$ and $K^{[4]}(\mathbf{0})$ in the asymptotic bias and variance of $T_{n,1}$ can be easily calculated. Additionally, considering (2.6), it is straightforward to prove that $\rho_c = 1/\lambda$. For simplicity, we also take $w(\mathbf{x}) = f(\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$. For each sample and in every scenario, the statistic $V^{-1/2}(T_{n,1} - b_{0\mathbf{H}})$ is computed.

Figure 2.3 shows density estimates of $V^{-1/2}(T_{n,1} - b_{0\mathbf{H}})$ (blue lines), computed with a Gaussian kernel and the rule-of-thumb bandwidth selector (Silverman, 1986), and the standard normal densities (red lines). The plot in the left panel corresponds to $n = 2500$ and the one in the right panel to $n = 10000$. When $n = 400$, the

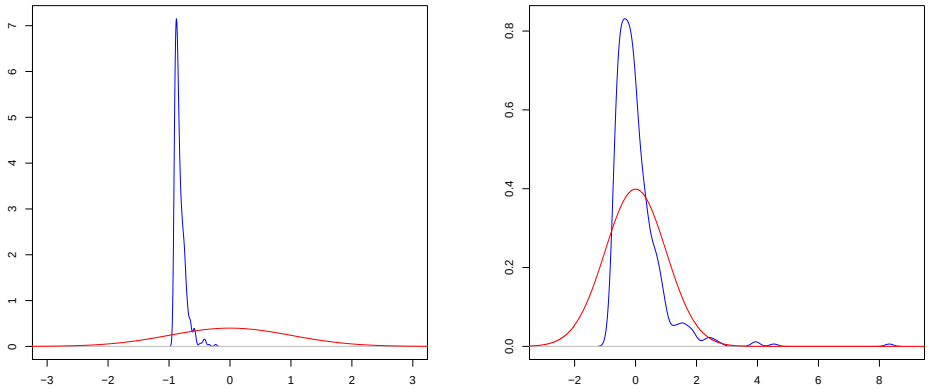

Figure 2.3: Density estimates of $V^{-1/2}(T_{n,1} - b_{0\mathbf{H}})$ (blue lines) and normal standard densities (red lines), considering $n = 2500$ (left panel) and $n = 10000$ (right panel).

asymptotic distribution of $V^{-1/2}(T_{n,1} - b_{0\mathbf{H}})$ is very far from the standard normal distribution and it is not shown here. Only when the sample size is very large, the sampling distribution of the test statistic seems to approximate reasonably well the Gaussian limit distribution. It is expected that this approximation will be better for larger sample sizes. That means that to obtain reliable results with the asymptotic distribution of the test, it would be necessary to consider a huge sample size (ignoring f and σ^2, which should be estimated). In this situation, the application of the test will take an enormous computing time.

In what follows, a detailed description of the different bootstrap proposals designed to perform the calibration of the test (namely PB, NPB and CNPB) will be presented. The main difference between the proposals is how the resampling residuals (required for mimicking the dependence structure) are computed. In PB, the residuals are obtained from the parametric regression estimator. Alternately, in NPB, the residuals are obtained from the nonparametric regression estimator (see González-Manteiga and Cao, 1993). In this way, the error variability could be reproduced consistently both under the null and the alternative hypotheses, increasing the power of the test. Finally, the CNPB procedure is a modification of the NPB, where the residuals are also obtained from the nonparametric regression estimator, but, additionally, the variability is estimated with an iterative algorithm to correct

the bias due to the use of the residuals (Fernández-Casal and Francisco-Fernández, 2014).

In order to describe the PB, NPB and CNPB resampling approaches, a generic bootstrap algorithm is firstly introduced. In what follows, no matter the method used, either parametric or nonparametric, $\hat{m}$ and $\hat{\boldsymbol{\Sigma}}$ denote the regression and the covariance matrix estimates, respectively.

Algorithm 1

1. Compute a parametric or a nonparametric regression estimator (described in Section 1.1.2), namely $\hat{m}(\mathbf{X}_i)$, $i = 1, \ldots, n$, depending if a parametric (PB) or a nonparametric (NPB or CNPB) bootstrap procedure is employed.
2. Obtain an estimated variance-covariance matrix $\hat{\boldsymbol{\Sigma}}$ of the residuals $\hat{\boldsymbol{\varepsilon}} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n)^{\mathrm{T}}$, where $\hat{\varepsilon}_i = Z_i - \hat{m}(\mathbf{X}_i)$, $i = 1, \ldots, n$.
3. Find the matrix $\mathbf{L}$, such that $\hat{\boldsymbol{\Sigma}} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, using Cholesky decomposition.
4. Compute the *independent* variables, $\mathbf{e} = (e_1, \ldots, e_n)^{\mathrm{T}}$, given by $\mathbf{e} = \mathbf{L}^{-1}\hat{\boldsymbol{\varepsilon}}$.
5. The previous *independent* variables are centered and an *independent* bootstrap sample of size n, denoted by $\mathbf{e}^* = (e_1^*, \ldots, e_n^*)^{\mathrm{T}}$, is obtained.
6. The bootstrap errors $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)^{\mathrm{T}}$ are computed as $\boldsymbol{\varepsilon}^* = \mathbf{L}\mathbf{e}^*$, and the bootstrap samples are $\{(\mathbf{X}_i, Z_i^*)\}_{i=1}^n$ with $Z_i^* = m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) + \varepsilon_i^*$, being $m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)$ the parametric regression estimator.
7. Using the bootstrap sample $\{(\mathbf{X}_i, Z_i^*)\}_{i=1}^n$, the bootstrap test statistic $T_{n,p}^*$ is computed as in (2.1).
8. Repeat Steps 5-7 a large number of times B.

The empirical distribution of the B bootstrap test statistics can be employed to approximate the finite sample distribution of the test statistic $T_{n,p}$ under the null hypothesis. Thus, denoting by $\{T_{n,1}^*, \cdots, T_{n,B}^*\}$ the sample of the B bootstrap test statistics, and defining its $(1-\alpha)$ quantile t_α^*, the null hypothesis in (1.34) will be rejected if $T_{n,p} > t_\alpha^*$. Additionally, the p-value of the test statistic can be approximated by:

$$p\text{-value} = \frac{1}{B}\sum_{b=1}^{B} \mathbb{I}_{\{T_{n,b}^* > T_{n,p}\}}. \tag{2.7}$$

Some steps of the Algorithm 1 are discussed below for PB, NPB and NCPB

methods. The main differences between the procedures are highlighted.

2.3.1 Parametric residual bootstrap (PB)

The PB extends to the case of spatial data the parametric residual bootstrap discussed in Vilar-Fernández and González-Manteiga (1996). In Step 1 of Algorithm 1, the regression function is estimated parametrically, employing the iterative least squares estimator constructed using (1.27). In Step 2, from the parametric residuals, the covariance matrix is also computed using a parametric approach (see Section 1.1.2 for further details). Notice that if the regression function and the semivariogram belong to the assumed parametric families, then this procedure should provide good results. However, a drawback of this procedure is the misspecification problem that may affect the regression and variance estimation. Moreover, as it was pointed out in Section 2.1, the direct use of the residuals introduces a bias in the estimation of the variability of the process in Step 2 of Algorithm 1.

2.3.2 Nonparametric residual bootstrap (NPB)

The NPB tries to avoid the misspecification problems mentioned in the previous section by using more flexible regression and dependence estimation methods than those employed in PB. In Step 1 of the bootstrap Algorithm 1, to increase the power of the test, following González-Manteiga and Cao (1993), the Nadaraya–Watson or the local linear estimators given in (1.6) or (1.9), respectively, are employed. In addition, in Step 2, a flexible procedure is considered to estimate the covariance matrix. The Shapiro–Botha variogram approach (Shapiro and Botha, 1991), combined with a nonparametric kernel semivariogram pilot estimation provides an efficient variogram estimator, which is used to approximate the corresponding covariance matrix. For more details see Section 1.1.2.

2.3.3 Corrected nonparametric residual bootstrap (CNPB)

As it was pointed out before, no matter the methodology used to remove the first-order structure in Step 2, either parametric or nonparametric, the direct use of the residuals in the variogram estimation introduces a bias in the approximation of the

process variability. The CNPB procedure is a modification of the previous NPB approach, considering a procedure to correct the resulting bias in the nonparametric estimator of the variogram. In the geostatistical framework, more accurate results have been obtained using this technique (Castillo-Páez et al., 2019). Specifically, the following adjustments are performed in the previous generic bootstrap Algorithm 1. In Step 1, the regression function is estimated using the Nadaraya–Watson or the local linear estimators given in (1.6) or in (1.9), respectively. In Step 2, from the corresponding nonparametric residuals, the dependence structure is estimated nonparametrically with an iterative algorithm to correct the bias (Fernández-Casal and Francisco-Fernández, 2014). Moreover, two additional steps are included after Step 2 and 3, which are denoted by 2* and 3*:

2*. Obtain a bias-corrected estimate of the variogram, using the residuals obtained from the nonparametric fit (see Fernández-Casal and Francisco-Fernández, 2014, for an exhaustive description of the algorithm) and calculate the corresponding (estimated) covariance matrix $\tilde{\mathbf{\Sigma}}$ of the errors.

3*. Find the matrix $\tilde{\mathbf{L}}$, such that $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathrm{T}}$, using Cholesky decomposition. $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^{\mathrm{T}}$.

In this situation, for the CNPB method, Step 6 in the Algorithm 1 needs to be modified as follows:

6. The bootstrap errors $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \ldots, \varepsilon_n^*)^{\mathrm{T}}$ are $\boldsymbol{\varepsilon}^* = \tilde{\mathbf{L}}\mathbf{e}^*$, and the bootstrap samples are $Z_i^* = m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) + \varepsilon_i^*$, where $m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)$ was computed using the procedure described in Section 1.1.2.

2.4 Simulation study

In this section, the practical performance of the proposed test statistic is analyzed through a simulation study comparing the different bootstrap procedures described in Section 2.3.

Three regression models are considered. In the first one, the parametric regression family $\mathcal{M}_{1,\beta}$, given in (2.4), is assumed for the null hypothesis. In this case, the

chosen regression functions have the expression:

$$m_1(\mathbf{X}) = X_1 + X_2 + c(X_2 + 1)^3, \quad \mathbf{X} = (X_1, X_2). \tag{2.8}$$

In the second one, the parametric regression family $\mathcal{M}_{2,\beta} = \{\beta_0 + \beta_1(X_1 - 0.5)^3, \beta_0, \beta_1 \in \mathbb{R}\}$ is supposed for the null hypothesis. In this case, the chosen regression functions have the expression:

$$m_2(\mathbf{X}) = 2.5 + 4(X_1 - 0.5)^3 + c\sin(2\pi X_2), \quad \mathbf{X} = (X_1, X_2). \tag{2.9}$$

The third parametric regression family considered for the null hypothesis is $\mathcal{M}_{3,\beta} = \{\beta_0 + \beta_1 \cos(\pi X_1), \beta_0, \beta_1 \in \mathbb{R}\}$, and the regression functions used have the expression:

$$m_3(\mathbf{X}) = 1 + 2\cos(\pi X_1) + c\sin(2\pi X_2), \quad \mathbf{X} = (X_1, X_2). \tag{2.10}$$

In all cases, the parameter c controls whether the null ($c = 0$) or the alternative ($c \neq 0$) hypotheses hold. For different values of this parameter ($c = 0, 0.5, 1$), 500 samples of sizes n ($n = 100, 225$ and 400) are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, following model (1.1), with regression functions (2.8), (2.9) or (2.10), and random errors ε_i normally distributed with zero mean and isotropic exponential covariogram:

$$\mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j) = c_e \left[\exp(-\|\mathbf{X}_i - \mathbf{X}_j\|/a_e)\right], \quad \|\mathbf{X}_i - \mathbf{X}_j\| \neq 0, \tag{2.11}$$

where c_e is the partial sill and a_e is the practical range, while the variance of the errors is $\sigma^2 = c_0 + c_e$, being c_0 the nugget effect. Different degrees of spatial dependence were studied, considering values of $a_e = 0.3, 0.6$ and 0.9, $\sigma^2 = 0.16, 0.32$ and 0.32, and nugget values of $0\%, 25\%$ and 50% of σ^2.

The behavior of the test statistic given in (2.1) was analyzed in the different scenarios. The parametric fit used to construct (2.1) was computed using the iterative least squares procedure described in Section 1.1.2. The nonparametric fit was obtained using the multivariate Nadaraya–Watson or the local linear estimators, given in (1.6) or (1.9), respectively, with a multiplicative triweight kernel. The bandwidth selection problem was addressed by employing the same procedure as that

used in Härdle and Mammen (1993), Alcalá et al. (1999), or Opsomer and Francisco-Fernández (2010), among others, analyzing the performance of the test statistic $T_{n,p}$ in (2.1) for a range of bandwidths. Initially, to simplify the calculations, the bandwidth matrix was restricted to a diagonal matrix with both equal elements (scalar matrix), $\mathbf{H} = \text{diag}(h, h)$, and different values of h in the interval $[0.25, 1.50]$ were chosen. The weight function employed in (2.1) to avoid the possible boundary effect (González-Manteiga and Cao, 1993) was $w(\mathbf{x}) = \mathbb{I}_{\{\mathbf{x}\in[1/\sqrt{n},1-1/\sqrt{n}]\times[1/\sqrt{n},1-1/\sqrt{n}]\}}$.

The bootstrap procedures described in Section 2.3 were applied using $B = 500$ replications. In Step 1 of Algorithm 1, for the nonparametric residual bootstrap procedures, NPB and CNPB, the multivariate Nadaraya–Watson or the local linear estimator (depending on whether $p = 0$ or $p = 1$, respectively) was computed using the optimal bandwidth that minimizes the MASE. Similar results were obtained when the MGCV bandwidth given in (1.33) is employed (Francisco-Fernandez and Opsomer, 2005). However, the use of the MASE bandwidth matrix reduces the computing time and avoids the effect of the bandwidth selection for the regression estimation on the results. Regarding the variogram, the (uncorrected) variogram estimates and the bias-corrected version were computed on a regular grid up to the 55% of the largest sample distance. In this case, the bandwidth matrices were selected applying the cross-validation relative squared error criterion. Figure 2.4 displays the theoretical semivariogram and the averaged values of the nonparametric semivariogram estimates considering $\mathcal{M}_{2,\beta}$ and $c = 0$, with $n = 400$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $c_0 = 0.04$, showing noticeable differences between both approaches.

The effect of the sample size as well as the impact of the spatial dependence degree on the behaviour of the test, under the null and under some alternative hypotheses, are analyzed below. In the different scenarios considered, a comparison of the proposed bootstrap procedures (PB, NPB and CNPB) is presented.

2.4.1 Sample size effect

In this section, the performance of the bootstrap procedures is analyzed for different sample sizes, under the null hypothesis and several alternatives. Proportions of rejections of the null hypothesis, for a significance level $\alpha = 0.05$, considering parameters $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ in model (2.11), and different sample sizes,

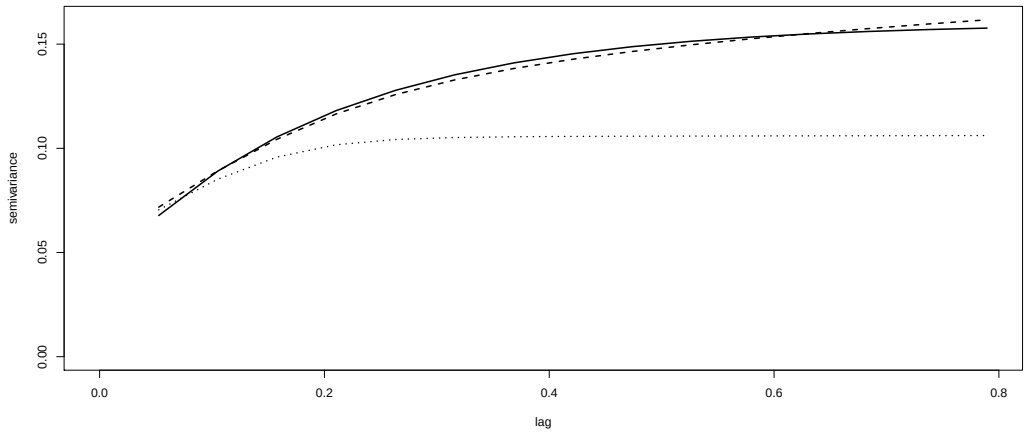

Figure 2.4: Theoretical error semivariogram (solid line) and averaged values of the semivariograms estimates obtained with the uncorrected residuals method (dotted line) and the corrected method (discontinuous line), considering $\mathcal{M}_{2,\beta}$, $n = 400$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $c_0 = 0.04$.

are computed. Tables 2.1, 2.2 and 2.3 show the results for the parametric families $\mathcal{M}_{1,\beta}$, $\mathcal{M}_{2,\beta}$ and $\mathcal{M}_{3,\beta}$, respectively. Under the null hypothesis ($c = 0$) and for the three parametric families, it can be observed that the test has a reasonable behavior when using PB and CNPB resampling methods. For both algorithms, the test preserves the nominal significance level of 5%, given that approximately 95% of the observed proportions of rejections under the null hypothesis to lie within the intervals $(0.007, 0.093)$, $(0.022, 0.078)$ and $(0.029, 0.071)$, when $n = 100, 225$ and 400, respectively. It should be noted that the proportions of rejections are affected by the value of h. For alternative assumptions ($c = 0.5$ and $c = 1$), in the case of $\mathcal{M}_{1,\beta}$, CNPB presents a slightly better performance than PB. In this situation, CNPB shows a decreasing power when the value of h increases. The opposite effect is observed when using PB. In the case of $\mathcal{M}_{2,\beta}$ and $\mathcal{M}_{3,\beta}$, under alternative hypotheses, the performance of PB is unsatisfactory. A much better behavior is observed for CNPB. Note that although it may seem that NPB presents a better behavior in terms of power, this is due to the underestimation of the variability, which induced really poor results under the null hypothesis. In general, no matter the parametric family considered, results obtained by the CNPB improve those achieved by PB and NPB,

for both null and alternative hypothesis, specially when the local linear estimator is employed, being, as expected, the power of the test larger when the value of c gets bigger.

2.4.2 Range of dependence effect

In this section, the performance of the bootstrap procedures is analyzed for different spatial dependence degrees ($a_e = 0.3$, $a_e = 0.6$ and $a_e = 0.9$). Values of $n = 400$, $\sigma^2 = 0.16$ and $c_0 = 0.04$ are considered. Figure 2.5 shows exponential variogram models with $\sigma^2 = 0.16$ and $c_0 = 0.04$, for $a_e = 0.3$ (black line), $a_e = 0.6$ (red line) and $a_e = 0.9$ (green line). Table 2.4 contains the proportions of rejections of the null hypothesis for $\alpha = 0.05$ employing $\mathcal{M}_{1,\beta}$. Notice that results for $a_e = 0.6$ have already been shown in Table 2.1, but for the sake of comparison they are also included in Table 2.4. If $c = 0$ (null hypothesis), the rejection proportions are similar to the theoretical level when using PB. Considering the local linear estimator and the NPB method, the behavior of the test is not very bad when the dependence structure is weak, but it gets worse as a_e is larger. Again, it can be observed that CNPB provides good results for the null and the alternative hypotheses. Results for the parametric family $\mathcal{M}_{2,\beta}$ are shown in Table 2.5. For larger values of the practical range a_e, the bandwidth values providing an effective calibration of the test should probably be larger. Regarding the PB approach, this resampling method works properly under the null hypothesis (for appropriate values of the bandwidth parameters h), but its performance under the alternatives is very poor. On the other hand, although the NPB method has a very high power, the proportions of rejections under the null hypothesis are very large. Similar conclusions at those given for $\mathcal{M}_{2,\beta}$ were obtained when the parametric family $\mathcal{M}_{3,\beta}$ was considered. Proportions of rejections of the null hypothesis for $\mathcal{M}_{3,\beta}$ are presented in Table 2.6.

2.4.3 Nugget effect

The performance of the proposed bootstrap procedures is now presented for different values of the nugget, 0%, 25% and 50% of σ^2. Figure 2.6 shows exponential variogram models with $\sigma^2 = 0.16$ and $a_e = 0.3$, for $c_0 = 0$ (black line), $c_0 = 0.04$ (red line) and

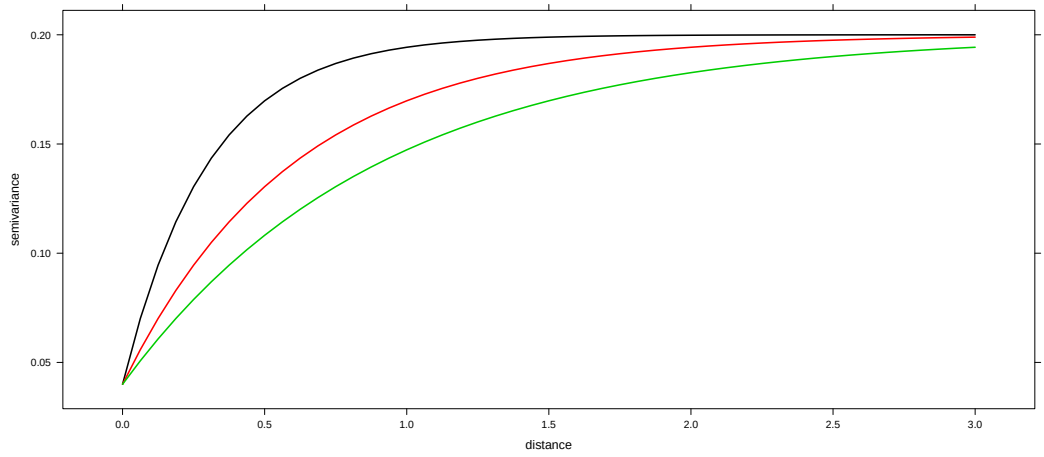

Figure 2.5: Exponential variogram models with $\sigma^2 = 0.16$ and $c_0 = 0.04$, for $a_e = 0.3$ (black line), $a_e = 0.6$ (red line) and $a_e = 0.9$ (green line)

$c_0 = 0.08$ (green line). Proportion of rejections of the null hypothesis are shown in Table 2.7, for $\alpha = 0.05$, considering the parametric family $\mathcal{M}_{1,\beta}$, $n = 400$, $\sigma^2 = 0.16$ and $a_e = 0.6$. Under the null hypothesis, both PB and CNPB, unlike NPB, provide reasonable results. Notice that in most cases, the local linear estimator provides better results than those obtained with the Nadaraya–Watson estimator. For instance, considering the Nadaraya–Watson estimator, the proportion of rejections of the null hypothesis are slightly larger or smaller than the theoretical level considering PB or CNPB, respectively. The power of the test is a bit larger when CNPB is employed. Once again, notice that although NPB provides the best results in terms of power, its performance under the null hypothesis is really poor. Proportions of rejection of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ are included in Table 2.8. The best behavior is observed when CNPB is employed, showing a good performance for the null and the different alternative hypotheses, specially considering the local linear estimator. For larger values of the variogram at zero lag, smaller bandwidths must be taken to calibrate the test properly. On the other hand, no reliable results are obtained for NPB under the null and the alternative hypotheses. Finally, regarding PB, the power of the test is very small for this scenario. Result for the parametric family $\mathcal{M}_{3,\beta}$ are summarized in Table 2.9. Similar conclusions to those provided for $\mathcal{M}_{2,\beta}$ were obtained. Notice that for the PB method and considering an alternative hypothesis pretty far apart from the null, the power of the test is reasonable when

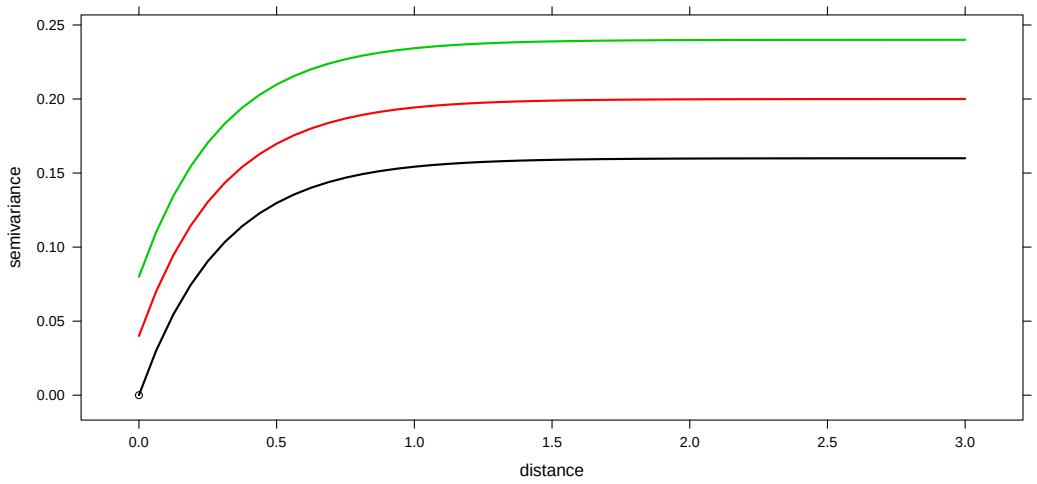

Figure 2.6: Exponential variogram models with $\sigma^2 = 0.16$ and $a_e = 0.3$, for $c_0 = 0$ (black line), $c_0 = 0.04$ (red line) and $c_0 = 0.08$ (green line)

the variogram at zero lag is equal to zero, but gets worse as c_0 is larger.

2.4.4 An experiment with non-Gaussian processes

The performance of the proposed test is analyzed when considering non-Gaussian errors in regression model (1.1). For this purpose, 500 samples are generated on a regular grid in the unit square $\mathcal{D} = [0,1]\times[0,1]$, following model (1.1), with regression function (2.9) and (non-Gaussian) random errors $\varepsilon_i = (\epsilon_i^2 - 1)/\sqrt{2}$, where ϵ_i are zero mean normally distributed random variables with covariance function equal to the square root of the one chosen for generating the Gaussian errors in the previous simulations. Notice that if the same simulation parameters are considered, this way of proceeding preserves the regression function and dependence structure (see Adler, 2010, Section 7.1). Table 2.10 summarizes the proportions of rejections for $\alpha = 0.05$, employing the parametric family $\mathcal{M}_{2,\boldsymbol{\beta}}$ and considering different values of the range parameter a_e, with $n = 400$, $\sigma^2 = 0.16$ and $c_0 = 0.04$. It can be observed that CNPB is the only procedure which works properly under both the null and the alternative hypothesis. As expected, for stronger dependence (that is, larger values of a_e), larger values of the bandwidth h must be selected to obtain an effective calibration of the test. Regarding the sample size and nugget effect, similar conclusions to those given

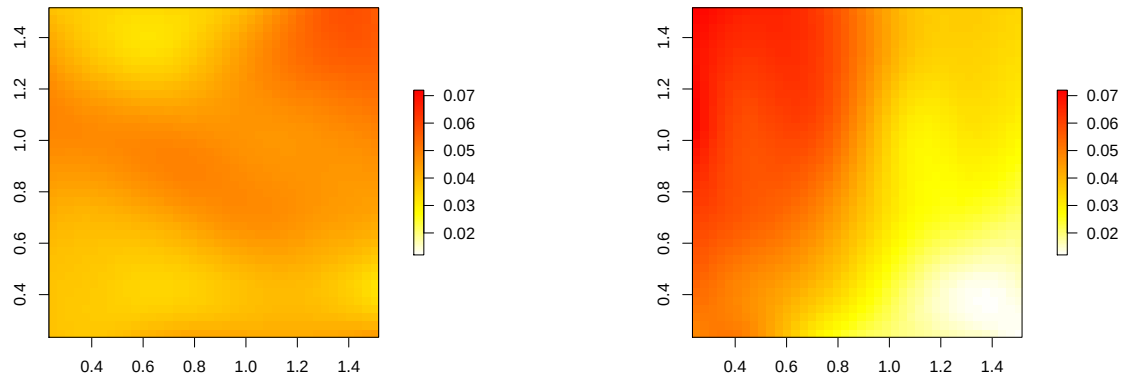

Figure 2.7: Proportions of rejections of the null hypothesis ($c = 0$), for $\alpha = 0.05$, considering the parametric family $\mathcal{M}_{2,\beta}$, $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$ and $n = 400$, using PB (left) and CNPB (right), for several values of h_1 and h_2.

for Gaussian errors were obtained when non-Gaussian errors were considered. For the sake of brevity, these results are not shown here. Note that no major differences (in terms of performance of the test) have been found if Gaussian or non-Gaussian errors are drawn in regression model (1.1).

2.4.5 More general bandwidth matrices

The impact on the performance of the test when more general bandwidth matrices are employed is analyzed in this section. Specifically, diagonal bandwidths with possible different elements, $\mathbf{H} = \text{diag}(h_1, h_2)$, are considered. Only some representative results employing the parametric family $\mathcal{M}_{2,\beta}$ are shown here. Values of $n = 400$, $\sigma^2 = 0.16$, $c_0 = 0.04$ and $a_e = 0.6$ were fixed. Proportions of rejections (under the null hypothesis, $c = 0$) for different combinations of h_1 and h_2, and considering $\alpha = 0.05$, are plotted in Figure 2.7. Left panel of Figure 2.7 shows the results for PB and right panel for CNPB. Proportions of rejections for NPB are omitted due to its deficient calibration. For this scenario, it can be observed that for PB there are not relevant differences in terms of rejection proportions if $\mathbf{H} = \text{diag}(h, h)$ or $\mathbf{H} = \text{diag}(h_1, h_2)$ (with $h_1 \neq h_2$) are considered. Regarding CNPB, the use of a more general bandwidth matrix does not provide better results with respect to using scalar bandwidth matrices. On the other hand, although it is omitted here, similar conclusions can be obtained for alternative hypotheses ($c \neq 0$).

Estimator	c	n	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	100	PB	0.134	0.116	0.114	0.106	0.082	0.088
			NPB	0.624	0.390	0.236	0.152	0.138	0.128
			CNPB	0.040	0.042	0.038	0.036	0.036	0.036
		225	PB	0.132	0.130	0.142	0.126	0.116	0.110
			NPB	0.596	0.404	0.244	0.182	0.162	0.154
			CNPB	0.012	0.014	0.016	0.016	0.020	0.020
		400	PB	0.074	0.088	0.108	0.112	0.096	0.094
			NPB	0.560	0.386	0.272	0.186	0.168	0.160
			CNPB	0.014	0.016	0.018	0.020	0.020	0.020
	0.5	100	PB	0.342	0.392	0.442	0.468	0.466	0.464
			NPB	0.990	0.946	0.896	0.848	0.826	0.808
			CNPB	0.460	0.448	0.402	0.390	0.370	0.360
		225	PB	0.306	0.382	0.464	0.502	0.510	0.520
			NPB	0.992	0.932	0.870	0.826	0.800	0.784
			CNPB	0.390	0.354	0.350	0.342	0.336	0.330
		400	PB	0.268	0.366	0.444	0.482	0.494	0.496
			NPB	0.980	0.954	0.898	0.866	0.834	0.826
			CNPB	0.370	0.326	0.336	0.336	0.328	0.334
	1	100	PB	0.898	0.922	0.930	0.946	0.948	0.956
			NPB	1.000	1.000	1.000	1.000	0.998	0.998
			CNPB	0.970	0.958	0.950	0.946	0.938	0.936
		225	PB	0.908	0.926	0.946	0.956	0.962	0.968
			NPB	1.000	1.000	0.998	0.998	0.998	0.996
			CNPB	0.970	0.968	0.950	0.942	0.934	0.934
		400	PB	0.904	0.934	0.956	0.966	0.970	0.970
			NPB	1.000	1.000	0.998	0.998	0.998	0.998
			CNPB	0.952	0.954	0.954	0.954	0.952	0.952
Local Linear	0	100	PB	0.132	0.114	0.100	0.088	0.076	0.068
			NPB	0.340	0.280	0.238	0.216	0.194	0.182
			CNPB	0.044	0.044	0.046	0.048	0.044	0.042
		225	PB	0.130	0.132	0.124	0.108	0.088	0.086
			NPB	0.284	0.276	0.250	0.224	0.214	0.192
			CNPB	0.034	0.032	0.038	0.038	0.038	0.038
		400	PB	0.074	0.090	0.078	0.064	0.064	0.068
			NPB	0.256	0.276	0.260	0.230	0.220	0.206
			CNPB	0.032	0.036	0.032	0.034	0.038	0.040
	0.5	100	PB	0.342	0.402	0.380	0.344	0.350	0.362
			NPB	0.982	0.968	0.946	0.920	0.888	0.854
			CNPB	0.638	0.642	0.582	0.524	0.496	0.498
		225	PB	0.310	0.362	0.322	0.298	0.322	0.364
			NPB	0.988	0.970	0.944	0.904	0.864	0.850
			CNPB	0.622	0.630	0.512	0.468	0.434	0.434
		400	PB	0.284	0.344	0.308	0.284	0.316	0.338
			NPB	0.984	0.966	0.934	0.912	0.874	0.856
			CNPB	0.596	0.582	0.502	0.446	0.430	0.444
	1	100	PB	0.898	0.938	0.946	0.930	0.930	0.936
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.992	0.998	0.992	0.984	0.980	0.980
		225	PB	0.914	0.942	0.938	0.930	0.934	0.942
			NPB	1.000	1.000	1.000	1.000	0.998	0.998
			CNPB	0.992	0.992	0.988	0.982	0.978	0.978
		400	PB	0.926	0.956	0.946	0.936	0.942	0.946
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.982	0.986	0.974	0.968	0.970	0.976

Table 2.1: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{1,\beta}$ with different sample sizes. Gaussian errors with $c_0 = 0.04$, $\sigma^2 = 0.16$, $a_e = 0.6$. Significance level: $\alpha = 0.05$.

Estimator	c	n	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	100	PB	0.056	0.042	0.054	0.066	0.070	0.068
			NPB	0.340	0.216	0.170	0.142	0.102	0.088
			CNPB	0.064	0.050	0.040	0.028	0.018	0.018
		225	PB	0.070	0.068	0.060	0.070	0.082	0.082
			NPB	0.268	0.192	0.176	0.144	0.116	0.108
			CNPB	0.078	0.058	0.046	0.042	0.030	0.028
		400	PB	0.038	0.036	0.046	0.046	0.052	0.056
			NPB	0.270	0.182	0.152	0.134	0.114	0.098
			CNPB	0.048	0.048	0.048	0.034	0.034	0.032
	0.5	100	PB	0.018	0.012	0.020	0.046	0.066	0.074
			NPB	1.000	0.996	0.978	0.960	0.898	0.818
			CNPB	0.740	0.574	0.380	0.198	0.076	0.050
		225	PB	0.004	0.004	0.016	0.036	0.058	0.070
			NPB	1.000	0.994	0.976	0.938	0.846	0.734
			CNPB	0.692	0.550	0.398	0.218	0.102	0.056
		400	PB	0.000	0.002	0.006	0.022	0.026	0.036
			NPB	1.000	1.000	0.984	0.940	0.852	0.716
			CNPB	0.384	0.316	0.208	0.076	0.034	0.028
	1	100	PB	0.056	0.010	0.018	0.038	0.074	0.110
			NPB	1.000	1.000	1.000	1.000	0.994	0.982
			CNPB	0.996	0.972	0.894	0.584	0.294	0.146
		225	PB	0.002	0.002	0.010	0.030	0.078	0.098
			NPB	1.000	1.000	1.000	1.000	0.994	0.962
			CNPB	0.990	0.938	0.848	0.564	0.296	0.126
		400	PB	0.000	0.000	0.000	0.024	0.064	0.098
			NPB	1.000	1.000	1.000	1.000	1.000	0.968
			CNPB	0.990	0.944	0.824	0.578	0.304	0.170
Local Linear	0	100	PB	0.056	0.048	0.056	0.056	0.052	0.050
			NPB	0.838	0.720	0.656	0.614	0.578	0.562
			CNPB	0.046	0.034	0.026	0.012	0.014	0.014
		225	PB	0.070	0.072	0.056	0.052	0.052	0.052
			NPB	0.866	0.756	0.676	0.640	0.610	0.604
			CNPB	0.036	0.040	0.032	0.032	0.022	0.022
		400	PB	0.038	0.046	0.048	0.042	0.042	0.040
			NPB	0.874	0.772	0.684	0.626	0.600	0.582
			CNPB	0.084	0.082	0.090	0.090	0.086	0.084
	0.5	100	PB	0.018	0.006	0.006	0.004	0.004	0.004
			NPB	1.000	0.998	0.988	0.986	0.980	0.976
			CNPB	0.540	0.350	0.170	0.054	0.024	0.016
		225	PB	0.004	0.004	0.004	0.004	0.004	0.004
			NPB	1.000	1.000	0.994	0.992	0.990	0.988
			CNPB	0.450	0.338	0.210	0.084	0.032	0.024
		400	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	0.996	0.996	0.992	0.992
			CNPB	0.620	0.408	0.392	0.444	0.436	0.424
	1	100	PB	0.056	0.004	0.004	0.004	0.004	0.004
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.988	0.954	0.800	0.400	0.152	0.048
		225	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.996	0.986	0.988	0.988	0.982
		400	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.990	0.968	0.984	0.984	0.972

Table 2.2: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different sample sizes. Gaussian errors with $c_0 = 0.04$, $\sigma^2 = 0.16$ and $a_e = 0.6$. Significance level: $\alpha = 0.05$.

Estimator	c	n	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	100	PB	0.076	0.064	0.064	0.076	0.082	0.080
			NPB	0.268	0.134	0.106	0.100	0.090	0.086
			CNPB	0.024	0.016	0.018	0.022	0.020	0.020
		225	PB	0.078	0.074	0.064	0.074	0.082	0.076
			NPB	0.202	0.092	0.086	0.084	0.080	0.074
			CNPB	0.012	0.012	0.020	0.020	0.018	0.020
		400	PB	0.036	0.040	0.044	0.054	0.056	0.058
			NPB	0.192	0.124	0.108	0.100	0.096	0.092
			CNPB	0.018	0.012	0.018	0.026	0.026	0.026
	0.5	100	PB	0.096	0.078	0.080	0.080	0.080	0.084
			NPB	1.000	0.990	0.966	0.902	0.822	0.682
			CNPB	0.478	0.304	0.144	0.052	0.040	0.034
		225	PB	0.048	0.052	0.062	0.066	0.066	0.066
			NPB	1.000	0.986	0.956	0.884	0.780	0.632
			CNPB	0.478	0.342	0.202	0.076	0.034	0.030
		400	PB	0.024	0.030	0.036	0.044	0.052	0.052
			NPB	1.000	0.996	0.972	0.926	0.824	0.658
			CNPB	0.438	0.334	0.194	0.080	0.034	0.026
	1	100	PB	0.368	0.142	0.112	0.104	0.106	0.108
			NPB	1.000	1.000	1.000	1.000	0.992	0.958
			CNPB	0.988	0.932	0.718	0.312	0.076	0.040
		225	PB	0.078	0.078	0.098	0.100	0.106	0.112
			NPB	1.000	1.000	1.000	1.000	0.996	0.950
			CNPB	0.994	0.938	0.788	0.412	0.168	0.072
		400	PB	0.024	0.032	0.046	0.062	0.076	0.090
			NPB	1.000	1.000	1.000	1.000	0.998	0.952
			CNPB	0.994	0.970	0.796	0.422	0.150	0.070
Local Linear	0	100	PB	0.076	0.074	0.072	0.068	0.062	0.060
			NPB	0.704	0.586	0.492	0.424	0.388	0.372
			CNPB	0.046	0.044	0.040	0.044	0.042	0.040
		225	PB	0.078	0.078	0.078	0.054	0.048	0.046
			NPB	0.782	0.620	0.486	0.432	0.396	0.358
			CNPB	0.024	0.034	0.040	0.036	0.034	0.034
		400	PB	0.034	0.038	0.044	0.044	0.044	0.040
			NPB	0.746	0.632	0.544	0.456	0.428	0.406
			CNPB	0.030	0.032	0.028	0.026	0.026	0.024
	0.5	100	PB	0.134	0.110	0.100	0.100	0.086	0.074
			NPB	0.996	0.990	0.970	0.968	0.960	0.940
			CNPB	0.288	0.156	0.146	0.158	0.150	0.142
		225	PB	0.128	0.136	0.132	0.118	0.092	0.098
			NPB	1.000	0.994	0.986	0.980	0.964	0.952
			CNPB	0.384	0.244	0.250	0.298	0.296	0.284
		400	PB	0.078	0.086	0.084	0.076	0.072	0.072
			NPB	1.000	0.998	0.992	0.984	0.980	0.976
			CNPB	0.354	0.232	0.238	0.306	0.302	0.290
	1	100	PB	0.358	0.080	0.058	0.058	0.058	0.058
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.988	0.834	0.732	0.742	0.724	0.702
		225	PB	0.066	0.020	0.018	0.024	0.028	0.026
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.960	0.940	0.980	0.960	0.960
		400	PB	0.018	0.010	0.010	0.012	0.012	0.012
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.958	0.938	0.966	0.964	0.964

Table 2.3: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{3,\beta}$ with different sample sizes. Gaussian errors with $c_0 = 0.04$, $\sigma^2 = 0.16$ and $a_e = 0.6$. Significance level: $\alpha = 0.05$.

Estimator	c	a_e	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0.3	PB	0.084	0.084	0.088	0.084	0.070	0.058
			NPB	0.554	0.300	0.188	0.118	0.098	0.092
			CNPB	0.020	0.020	0.020	0.022	0.022	0.022
		0.6	PB	0.074	0.088	0.108	0.112	0.096	0.094
			NPB	0.560	0.386	0.272	0.186	0.168	0.160
			CNPB	0.014	0.016	0.018	0.020	0.020	0.020
		0.9	PB	0.048	0.062	0.084	0.092	0.094	0.106
			NPB	0.570	0.430	0.302	0.216	0.200	0.194
			CNPB	0.010	0.014	0.022	0.022	0.022	0.022
	0.5	0.3	PB	0.178	0.294	0.424	0.488	0.520	0.532
			NPB	0.936	0.786	0.652	0.596	0.562	0.552
			CNPB	0.370	0.300	0.324	0.342	0.332	0.330
		0.6	PB	0.268	0.366	0.444	0.482	0.494	0.496
			NPB	0.980	0.954	0.898	0.866	0.834	0.826
			CNPB	0.370	0.326	0.336	0.336	0.328	0.334
		0.9	PB	0.342	0.410	0.486	0.514	0.530	0.538
			NPB	0.984	0.942	0.896	0.866	0.838	0.828
			CNPB	0.392	0.368	0.372	0.374	0.380	0.372
	1	0.3	PB	0.904	0.940	0.960	0.966	0.978	0.978
			NPB	0.998	0.996	0.988	0.978	0.970	0.968
			CNPB	0.988	0.986	0.980	0.984	0.984	0.986
		0.6	PB	0.904	0.934	0.956	0.966	0.970	0.970
			NPB	1.000	1.000	0.998	0.998	0.998	0.998
			CNPB	0.952	0.954	0.954	0.954	0.952	0.952
		0.9	PB	0.940	0.960	0.972	0.974	0.978	0.978
			NPB	1.000	1.000	0.998	0.998	0.998	0.998
			CNPB	0.952	0.958	0.954	0.952	0.954	0.956
Local Linear	0	0.3	PB	0.084	0.096	0.088	0.068	0.060	0.064
			NPB	0.096	0.112	0.126	0.112	0.104	0.090
			CNPB	0.018	0.018	0.026	0.022	0.024	0.024
		0.6	PB	0.074	0.090	0.078	0.064	0.064	0.068
			NPB	0.256	0.276	0.260	0.230	0.220	0.206
			CNPB	0.032	0.036	0.032	0.034	0.038	0.040
		0.9	PB	0.048	0.060	0.050	0.048	0.042	0.038
			NPB	0.330	0.364	0.312	0.292	0.296	0.292
			CNPB	0.026	0.026	0.028	0.036	0.046	0.054
	0.5	0.3	PB	0.188	0.272	0.206	0.180	0.208	0.242
			NPB	0.814	0.750	0.648	0.540	0.480	0.468
			CNPB	0.650	0.646	0.548	0.458	0.440	0.448
		0.6	PB	0.284	0.344	0.308	0.284	0.316	0.338
			NPB	0.984	0.966	0.934	0.912	0.874	0.856
			CNPB	0.596	0.582	0.502	0.446	0.430	0.444
		0.9	PB	0.352	0.410	0.368	0.354	0.378	0.412
			NPB	0.988	0.980	0.954	0.944	0.912	0.900
			CNPB	0.644	0.640	0.558	0.498	0.492	0.508
	1	0.3	PB	0.922	0.952	0.928	0.916	0.926	0.942
			NPB	0.998	0.996	0.994	0.984	0.964	0.960
			CNPB	0.996	0.996	0.996	0.992	0.988	0.990
		0.6	PB	0.926	0.956	0.946	0.936	0.942	0.946
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.982	0.986	0.974	0.968	0.970	0.976
		0.9	PB	0.954	0.974	0.966	0.964	0.966	0.972
			NPB	1.000	1.000	1.000	1.000	0.998	0.998
			CNPB	0.984	0.988	0.980	0.972	0.978	0.980

Table 2.4: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{1,\beta}$ with different range values. Gaussian errors with $n = 400$, $c_0 = 0.04$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	a_e	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0.3	PB	0.056	0.050	0.040	0.044	0.044	0.050
			NPB	0.158	0.090	0.076	0.062	0.050	0.052
			CNPB	0.030	0.024	0.018	0.012	0.010	0.010
		0.6	PB	0.038	0.036	0.046	0.046	0.052	0.056
			NPB	0.270	0.182	0.152	0.134	0.114	0.098
			CNPB	0.048	0.048	0.048	0.034	0.034	0.032
		0.9	PB	0.002	0.004	0.018	0.030	0.040	0.042
			NPB	0.328	0.266	0.232	0.186	0.162	0.144
			CNPB	0.070	0.070	0.068	0.060	0.048	0.048
	0.5	0.3	PB	0.000	0.002	0.006	0.022	0.026	0.036
			NPB	1.000	1.000	0.998	0.976	0.878	0.716
			CNPB	0.346	0.270	0.170	0.042	0.018	0.016
		0.6	PB	0.004	0.004	0.016	0.038	0.058	0.066
			NPB	1.000	1.000	0.984	0.940	0.852	0.716
			CNPB	0.384	0.316	0.208	0.076	0.034	0.028
		0.9	PB	0.008	0.010	0.020	0.048	0.062	0.072
			NPB	1.000	0.996	0.982	0.954	0.880	0.750
			CNPB	0.526	0.432	0.284	0.146	0.074	0.052
	1	0.3	PB	0.000	0.000	0.000	0.010	0.050	0.062
			NPB	1.000	1.000	1.000	1.000	1.000	0.986
			CNPB	1.000	0.990	0.906	0.648	0.380	0.194
		0.6	PB	0.000	0.000	0.000	0.024	0.064	0.098
			NPB	1.000	1.000	1.000	1.000	1.000	0.968
			CNPB	0.990	0.944	0.824	0.578	0.304	0.170
		0.9	PB	0.000	0.000	0.008	0.030	0.080	0.110
			NPB	1.000	1.000	1.000	1.000	1.000	0.968
			CNPB	1.000	0.996	0.988	0.992	0.988	0.984
Local Linear	0	0.3	PB	0.056	0.074	0.072	0.062	0.060	0.058
			NPB	0.702	0.476	0.400	0.366	0.350	0.326
			CNPB	0.024	0.020	0.012	0.006	0.006	0.006
		0.6	PB	0.038	0.046	0.048	0.042	0.042	0.040
			NPB	0.874	0.772	0.684	0.626	0.600	0.582
			CNPB	0.084	0.082	0.090	0.090	0.086	0.084
		0.9	PB	0.004	0.010	0.016	0.010	0.014	0.014
			NPB	0.922	0.852	0.792	0.748	0.722	0.708
			CNPB	0.052	0.060	0.060	0.050	0.040	0.038
	0.5	0.3	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.352	0.272	0.168	0.044	0.018	0.016
		0.6	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	0.996	0.996	0.992	0.992
			CNPB	0.620	0.408	0.392	0.444	0.436	0.424
		0.9	PB	0.004	0.002	0.002	0.002	0.002	0.002
			NPB	1.000	1.000	0.998	0.994	0.990	0.990
			CNPB	0.712	0.518	0.498	0.522	0.518	0.500
	1	0.3	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.994	0.928	0.762	0.456	0.188	0.074
		0.6	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.990	0.968	0.984	0.984	0.972
		0.9	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.994	0.986	0.992	0.988	0.982

Table 2.5: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different range values. Gaussian errors with $n = 400$, $c_0 = 0.04$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	a_e	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0.3	PB	0.070	0.068	0.048	0.042	0.038	0.044
			NPB	0.130	0.056	0.046	0.044	0.046	0.046
			CNPB	0.012	0.008	0.004	0.012	0.010	0.010
		0.6	PB	0.036	0.040	0.044	0.054	0.056	0.058
			NPB	0.192	0.124	0.108	0.100	0.096	0.092
			CNPB	0.018	0.012	0.018	0.026	0.026	0.026
		0.9	PB	0.022	0.024	0.034	0.054	0.060	0.058
			NPB	0.208	0.146	0.130	0.122	0.124	0.122
			CNPB	0.024	0.024	0.026	0.030	0.030	0.026
	0.5	0.3	PB	0.004	0.006	0.016	0.020	0.024	0.024
			NPB	1.000	1.000	0.996	0.976	0.856	0.656
			CNPB	0.400	0.320	0.162	0.038	0.020	0.016
		0.6	PB	0.024	0.030	0.036	0.044	0.052	0.052
			NPB	1.000	0.996	0.972	0.926	0.824	0.658
			CNPB	0.438	0.334	0.194	0.080	0.034	0.026
		0.9	PB	0.032	0.052	0.088	0.098	0.102	0.102
			NPB	1.000	0.994	0.970	0.920	0.848	0.686
			CNPB	0.536	0.412	0.254	0.118	0.060	0.054
	1	0.3	PB	0.010	0.014	0.018	0.034	0.048	0.054
			NPB	1.000	1.000	1.000	1.000	1.000	0.984
			CNPB	0.998	0.976	0.812	0.426	0.158	0.060
		0.6	PB	0.024	0.032	0.046	0.062	0.076	0.090
			NPB	1.000	1.000	1.000	1.000	0.998	0.952
			CNPB	0.994	0.970	0.796	0.422	0.150	0.070
		0.9	PB	0.034	0.050	0.064	0.080	0.102	0.112
			NPB	1.000	1.000	1.000	1.000	0.996	0.968
			CNPB	0.996	0.974	0.852	0.486	0.212	0.094
Local Linear	0	0.3	PB	0.070	0.084	0.078	0.066	0.060	0.056
			NPB	0.598	0.386	0.306	0.256	0.218	0.206
			CNPB	0.074	0.054	0.046	0.040	0.038	0.036
		0.6	PB	0.034	0.038	0.044	0.044	0.044	0.040
			NPB	0.746	0.632	0.544	0.456	0.428	0.406
			CNPB	0.148	0.146	0.142	0.128	0.124	0.120
		0.9	PB	0.022	0.026	0.028	0.026	0.032	0.034
			NPB	0.816	0.730	0.636	0.552	0.522	0.506
			CNPB	0.220	0.216	0.214	0.200	0.192	0.184
	0.5	0.3	PB	0.004	0.002	0.002	0.002	0.002	0.004
			NPB	1.000	0.998	0.996	0.996	0.996	0.994
			CNPB	0.312	0.200	0.220	0.310	0.316	0.302
		0.6	PB	0.022	0.008	0.008	0.018	0.018	0.018
			NPB	1.000	0.998	0.992	0.984	0.980	0.976
			CNPB	0.006	0.006	0.006	0.008	0.008	0.008
		0.9	PB	0.028	0.022	0.022	0.026	0.026	0.026
			NPB	1.000	0.998	0.988	0.982	0.972	0.970
			CNPB	0.424	0.304	0.310	0.340	0.344	0.332
	1	0.3	PB	0.006	0.002	0.004	0.004	0.006	0.006
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.950	0.946	0.964	0.964	0.960
		0.6	PB	0.018	0.010	0.010	0.012	0.012	0.012
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.958	0.938	0.966	0.964	0.964
		0.9	PB	0.028	0.016	0.018	0.022	0.022	0.024
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.998	0.966	0.960	0.970	0.966	0.964

Table 2.6: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{3,\beta}$ with different range values. Gaussian errors with $n = 400$, $c_0 = 0.04$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	c_0	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0%	PB	0.064	0.062	0.086	0.094	0.106	0.106
			NPB	0.562	0.412	0.288	0.206	0.188	0.178
			CNPB	0.014	0.014	0.014	0.020	0.018	0.018
		25%	PB	0.074	0.088	0.108	0.112	0.096	0.094
			NPB	0.558	0.388	0.276	0.186	0.166	0.160
			CNPB	0.014	0.016	0.018	0.020	0.020	0.020
		50%	PB	0.110	0.100	0.104	0.106	0.106	0.096
			NPB	0.648	0.398	0.268	0.166	0.138	0.132
			CNPB	0.038	0.032	0.028	0.028	0.030	0.030
	0.5	0%	PB	0.348	0.390	0.458	0.476	0.488	0.498
			NPB	0.974	0.930	0.860	0.808	0.790	0.780
			CNPB	0.308	0.282	0.302	0.304	0.306	0.304
		25%	PB	0.268	0.366	0.444	0.482	0.494	0.496
			NPB	0.980	0.954	0.898	0.866	0.834	0.826
			CNPB	0.370	0.326	0.336	0.336	0.328	0.334
		50%	PB	0.310	0.398	0.500	0.540	0.566	0.572
			NPB	0.994	0.976	0.944	0.900	0.886	0.868
			CNPB	0.472	0.444	0.418	0.432	0.430	0.430
	1	0%	PB	0.938	0.944	0.954	0.956	0.958	0.962
			NPB	1.000	1.000	0.996	0.994	0.996	0.996
			CNPB	0.918	0.912	0.912	0.906	0.906	0.904
		25%	PB	0.904	0.934	0.956	0.966	0.970	0.970
			NPB	1.000	1.000	0.998	0.998	0.998	0.998
			CNPB	0.952	0.954	0.954	0.954	0.952	0.952
		50%	PB	0.956	0.974	0.980	0.980	0.984	0.986
			NPB	1.000	1.000	1.000	0.998	0.998	0.998
			CNPB	0.990	0.984	0.988	0.986	0.984	0.984
Local Linear	0	0%	PB	0.080	0.078	0.056	0.044	0.052	0.070
			NPB	0.384	0.370	0.318	0.292	0.282	0.266
			CNPB	0.032	0.030	0.034	0.040	0.046	0.052
		25%	PB	0.074	0.090	0.078	0.064	0.064	0.068
			NPB	0.256	0.276	0.260	0.230	0.220	0.206
			CNPB	0.032	0.036	0.032	0.034	0.038	0.040
		50%	PB	0.112	0.110	0.100	0.080	0.078	0.066
			NPB	0.180	0.222	0.206	0.180	0.158	0.144
			CNPB	0.030	0.028	0.030	0.030	0.032	0.028
	0.5	0%	PB	0.370	0.398	0.336	0.302	0.332	0.366
			NPB	0.984	0.956	0.926	0.890	0.852	0.834
			CNPB	0.542	0.510	0.426	0.376	0.382	0.406
		25%	PB	0.284	0.344	0.308	0.284	0.316	0.338
			NPB	0.984	0.966	0.934	0.912	0.874	0.856
			CNPB	0.596	0.582	0.502	0.446	0.430	0.444
		50%	PB	0.316	0.388	0.340	0.328	0.346	0.374
			NPB	0.992	0.984	0.972	0.956	0.928	0.900
			CNPB	0.712	0.716	0.638	0.566	0.546	0.554
	1	0%	PB	0.952	0.968	0.954	0.934	0.934	0.940
			NPB	1.000	1.000	1.000	1.000	0.998	0.998
			CNPB	0.966	0.964	0.958	0.950	0.950	0.954
		25%	PB	0.926	0.956	0.946	0.936	0.942	0.946
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.982	0.986	0.974	0.968	0.970	0.976
		50%	PB	0.964	0.976	0.970	0.970	0.970	0.972
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.996	0.994	0.992	0.990	0.990

Table 2.7: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{1,\beta}$ with different nugget values. Gaussian errors with $n = 400$, $a_e = 0.6$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	c_0	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0%	PB	0.014	0.008	0.020	0.034	0.046	0.046
			NPB	0.338	0.230	0.182	0.154	0.132	0.110
			CNPB	0.048	0.048	0.042	0.034	0.032	0.032
		25%	PB	0.038	0.036	0.046	0.046	0.052	0.056
			NPB	0.270	0.182	0.152	0.134	0.114	0.098
			CNPB	0.048	0.048	0.048	0.034	0.034	0.032
		50%	PB	0.050	0.048	0.042	0.048	0.052	0.056
			NPB	0.254	0.172	0.144	0.116	0.092	0.082
			CNPB	0.050	0.050	0.046	0.036	0.030	0.028
	0.5	0%	PB	0.006	0.014	0.034	0.052	0.062	0.068
			NPB	1.000	0.990	0.972	0.912	0.834	0.686
			CNPB	0.304	0.248	0.154	0.048	0.032	0.024
		25%	PB	0.004	0.004	0.016	0.038	0.058	0.066
			NPB	1.000	1.000	0.984	0.940	0.852	0.716
			CNPB	0.640	0.534	0.384	0.206	0.096	0.058
		50%	PB	0.000	0.002	0.010	0.026	0.048	0.056
			NPB	1.000	1.000	0.998	0.976	0.890	0.776
			CNPB	0.880	0.780	0.652	0.468	0.264	0.164
	1	0%	PB	0.104	0.074	0.092	0.132	0.182	0.204
			NPB	1.000	1.000	1.000	1.000	0.992	0.922
			CNPB	0.926	0.814	0.658	0.334	0.150	0.056
		25%	PB	0.000	0.000	0.000	0.024	0.064	0.098
			NPB	1.000	1.000	1.000	1.000	1.000	0.968
			CNPB	0.990	0.944	0.824	0.578	0.304	0.170
		50%	PB	0.000	0.000	0.004	0.014	0.052	0.076
			NPB	1.000	1.000	1.000	1.000	1.000	0.996
			CNPB	1.000	1.000	0.970	0.840	0.576	0.378
Local Linear	0	0%	PB	0.020	0.034	0.034	0.030	0.026	0.030
			NPB	0.906	0.802	0.714	0.650	0.616	0.604
			CNPB	0.046	0.048	0.038	0.030	0.028	0.024
		25%	PB	0.038	0.046	0.048	0.042	0.042	0.040
			NPB	0.874	0.772	0.684	0.626	0.600	0.582
			CNPB	0.084	0.082	0.090	0.090	0.086	0.084
		50%	PB	0.050	0.050	0.054	0.048	0.042	0.040
			NPB	0.868	0.748	0.670	0.600	0.572	0.558
			CNPB	0.110	0.094	0.104	0.096	0.090	0.086
	0.5	0%	PB	0.004	0.004	0.004	0.004	0.004	0.004
			NPB	1.000	1.000	0.992	0.990	0.986	0.982
			CNPB	0.522	0.344	0.332	0.360	0.366	0.350
		25%	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	0.996	0.996	0.992	0.992
			CNPB	0.620	0.408	0.392	0.444	0.436	0.424
		50%	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.746	0.540	0.518	0.562	0.562	0.546
	1	0%	PB	0.064	0.012	0.008	0.012	0.016	0.014
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.964	0.944	0.962	0.956	0.944
		25%	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.990	0.968	0.984	0.984	0.972
		50%	PB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.992	0.984	0.990	0.986	0.986

Table 2.8: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different nugget values. Gaussian errors with $n = 400$, $a_e = 0.6$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	c_0	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0%	PB	0.032	0.020	0.040	0.054	0.062	0.064
			NPB	0.250	0.152	0.130	0.120	0.120	0.118
			CNPB	0.022	0.018	0.024	0.028	0.030	0.030
		25%	PB	0.036	0.040	0.044	0.054	0.056	0.058
			NPB	0.188	0.124	0.106	0.100	0.096	0.092
			CNPB	0.018	0.014	0.016	0.026	0.026	0.026
		50%	PB	0.054	0.052	0.056	0.052	0.050	0.052
			NPB	0.160	0.094	0.074	0.068	0.072	0.072
			CNPB	0.016	0.012	0.016	0.024	0.026	0.026
	0.5	0%	PB	0.052	0.054	0.058	0.062	0.064	0.064
			NPB	1.000	0.996	0.964	0.914	0.824	0.678
			CNPB	0.466	0.380	0.246	0.102	0.050	0.038
		25%	PB	0.024	0.030	0.036	0.044	0.052	0.052
			NPB	1.000	0.996	0.972	0.926	0.824	0.656
			CNPB	0.438	0.334	0.194	0.080	0.034	0.026
		50%	PB	0.010	0.026	0.038	0.040	0.044	0.052
			NPB	1.000	0.998	0.992	0.956	0.856	0.660
			CNPB	0.612	0.464	0.314	0.140	0.056	0.036
	1	0%	PB	0.434	0.336	0.270	0.236	0.220	0.216
			NPB	1.000	1.000	1.000	1.000	0.992	0.942
			CNPB	0.988	0.922	0.710	0.334	0.102	0.064
		25%	PB	0.024	0.032	0.046	0.062	0.076	0.090
			NPB	1.000	1.000	1.000	1.000	0.998	0.952
			CNPB	0.994	0.970	0.796	0.422	0.150	0.070
		50%	PB	0.024	0.040	0.052	0.070	0.082	0.090
			NPB	1.000	1.000	1.000	1.000	1.000	0.994
			CNPB	1.000	0.988	0.884	0.550	0.292	0.132
Local Linear	0	0%	PB	0.030	0.030	0.026	0.034	0.034	0.034
			NPB	0.802	0.668	0.560	0.486	0.456	0.432
			CNPB	0.042	0.044	0.040	0.036	0.036	0.036
		25%	PB	0.034	0.038	0.044	0.044	0.044	0.040
			NPB	0.748	0.630	0.544	0.456	0.426	0.406
			CNPB	0.030	0.032	0.028	0.026	0.026	0.024
		50%	PB	0.054	0.056	0.062	0.050	0.044	0.042
			NPB	0.734	0.602	0.492	0.416	0.382	0.368
			CNPB	0.036	0.036	0.036	0.036	0.028	0.030
	0.5	0%	PB	0.040	0.026	0.026	0.032	0.032	0.034
			NPB	1.000	0.998	0.990	0.976	0.968	0.964
			CNPB	0.306	0.218	0.230	0.266	0.272	0.272
		25%	PB	0.022	0.008	0.008	0.018	0.018	0.018
			NPB	1.000	0.998	0.992	0.984	0.980	0.976
			CNPB	0.006	0.006	0.006	0.008	0.008	0.008
		50%	PB	0.010	0.006	0.006	0.006	0.006	0.006
			NPB	1.000	0.998	0.994	0.994	0.992	0.988
			CNPB	0.494	0.330	0.342	0.414	0.420	0.402
	1	0%	PB	0.368	0.140	0.144	0.182	0.186	0.176
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.982	0.924	0.910	0.940	0.936	0.932
		25%	PB	0.018	0.010	0.010	0.012	0.012	0.012
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.958	0.938	0.966	0.964	0.964
		50%	PB	0.020	0.006	0.006	0.010	0.010	0.010
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.996	0.978	0.966	0.984	0.984	0.980

Table 2.9: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{3,\beta}$ with different nugget values. Gaussian errors with $n = 400$, $a_e = 0.6$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

Estimator	c	a_e	Method	$h = 0.25$	$h = 0.50$	$h = 0.75$	$h = 1.00$	$h = 1.25$	$h = 1.50$
Nadaraya–Watson	0	0.3	SPB	0.142	0.062	0.070	0.112	0.142	0.158
			NPB	0.120	0.072	0.066	0.050	0.042	0.042
			CNPB	0.020	0.022	0.010	0.008	0.008	0.008
		0.6	SPB	0.030	0.020	0.032	0.042	0.058	0.062
			NPB	0.228	0.160	0.142	0.120	0.096	0.084
			CNPB	0.048	0.044	0.036	0.032	0.030	0.030
		0.9	SPB	0.028	0.016	0.034	0.052	0.072	0.074
			NPB	0.302	0.238	0.198	0.172	0.146	0.134
			CNPB	0.062	0.062	0.060	0.046	0.042	0.040
	0.5	0.3	SPB	0.000	0.000	0.002	0.008	0.022	0.026
			NPB	1.000	1.000	1.000	0.996	0.928	0.734
			CNPB	0.360	0.254	0.116	0.016	0.004	0.004
		0.6	SPB	0.000	0.000	0.004	0.022	0.036	0.046
			NPB	1.000	1.000	1.000	0.982	0.898	0.716
			CNPB	0.354	0.276	0.132	0.028	0.014	0.008
		0.9	SPB	0.000	0.002	0.010	0.022	0.036	0.040
			NPB	1.000	1.000	0.998	0.982	0.906	0.750
			CNPB	0.510	0.400	0.228	0.068	0.022	0.012
	1	0.3	SPB	0.000	0.000	0.000	0.004	0.032	0.052
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	0.996	0.888	0.494	0.170	0.050
		0.6	SPB	0.008	0.008	0.014	0.034	0.072	0.100
			NPB	1.000	1.000	1.000	1.000	1.000	0.988
			CNPB	1.000	0.980	0.810	0.416	0.126	0.028
		0.9	SPB	0.036	0.024	0.032	0.046	0.084	0.106
			NPB	1.000	1.000	1.000	1.000	1.000	0.994
			CNPB	1.000	0.990	0.878	0.502	0.164	0.044
Local Linear	0	0.3	SPB	0.142	0.062	0.070	0.112	0.142	0.158
			NPB	0.120	0.072	0.066	0.050	0.042	0.042
			CNPB	0.020	0.022	0.010	0.008	0.008	0.008
		0.6	SPB	0.030	0.020	0.032	0.042	0.058	0.062
			NPB	0.228	0.160	0.142	0.120	0.096	0.084
			CNPB	0.048	0.044	0.036	0.032	0.030	0.030
		0.9	SPB	0.032	0.018	0.018	0.014	0.018	0.020
			NPB	0.302	0.238	0.198	0.172	0.146	0.134
			CNPB	0.062	0.062	0.060	0.046	0.042	0.040
	0.5	0.3	SPB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.696	0.412	0.400	0.492	0.500	0.480
		0.6	SPB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	0.676	0.382	0.366	0.444	0.446	0.420
		0.9	SPB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	0.998	0.998	0.998	0.998
			CNPB	0.774	0.484	0.460	0.532	0.534	0.514
	1	0.3	SPB	0.000	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	1.000	1.000	1.000	1.000	1.000
		0.6	SPB	0.006	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	1.000	1.000	1.000	1.000	1.000
		0.9	SPB	0.020	0.000	0.000	0.000	0.000	0.000
			NPB	1.000	1.000	1.000	1.000	1.000	1.000
			CNPB	1.000	1.000	1.000	1.000	1.000	1.000

Table 2.10: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different range values. Non-Gaussian errors with $n = 400$, $c_0 = 0.04$ and $\sigma^2 = 0.16$. Significance level: $\alpha = 0.05$.

2.5 Real data illustration

In order to illustrate the performance in practice of the test statistic $T_{n,p}$, given in (2.1), the Wolfcamp aquifer dataset, briefly introduced in Section 2.1, is considered. These data were reported and geostatistically analyzed in Harper and Furr (1986) and Cressie (1993), and are available in the `R` package `npsp` (Fernández-Casal, 2019). Based on the simulation study, where both $T_{n,0}$ and $T_{n,1}$ presented a similar performance, just results using the local linear estimator ($T_{n,1}$) are provided in this section. Moreover, taking into account that the results obtained by the CNPB were clearly better than those achieved by PB and NPB in the simulations, only the CNPB resampling approach will be used to calibrate the test.

The Deaf Smith County (Texas, bordering New Mexico) was selected as an alternate site for a possible nuclear waste disposal repository in the 1980s. This site was later dropped on grounds of contamination of the aquifer, the source of much of the water supply for west Texas. In a study conducted by the U.S. Department of Energy, piezometric-head levels were obtained irregularly at 85 locations, shown in Figure 2.1 (included in Section 2.1), by drilling a narrow pipe through the aquifer (see Harper and Furr, 1986). With higher values generally in the lower left (southwest) and lower values in the upper right (northwest), the groundwater gradient would cause water to flow in a northeasterly direction from the repository in Deaf Smith County toward Amarillo in lower Potter County.

Figure 2.2 (shown in Section 2.1) displayed the 3-dimensional scatterplot of the piezometric heads levels (feet above sea level) against over the coordinates (miles, from a reference point). This plot evidenced a clear downwards trend from southwest to northeast. Cressie (1993) used the median polish approach to model this trend, whereas Harper and Furr (1986) considered a linear trend surface, that is, a linear regression model on latitude and longitude.

In this section, in order to check if a linear model is plausible, the test $T_{n,1}$, using the CNPB procedure with $B = 500$ replications, was applied considering the linear parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}$, given in (2.4), as the null hypothesis. In this case, X_1 and X_2 are the spatial coordinates of the points where the process is observed. It should be noted that the (nonparametric) detrended data were also tested for isotropy and stationarity, following the proposals by Bowman and Crujeiras (2013), obtaining

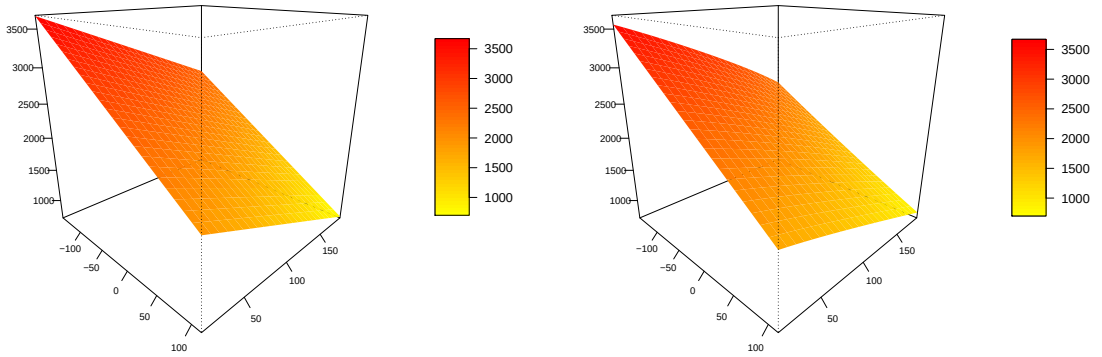

Figure 2.8: Smooth version of the parametric fit (left) and nonparametric estimator of the regression (right) using the MGCV bandwidth for the Wolfcamp Aquifer.

p-values of 0.838 for isotropy and 0.031 for stationarity.

To apply the test (2.1), the parametric fit was carried out using the iterative least squares estimator described in Section 1.1.2, assuming a linear regression model. After analyzing the initial residuals obtained by least squares regression, a spherical correlation model (as it was suggested by Harper and Furr, 1986) was considered to estimate the variance-covariance matrix of the errors, needed to obtain a feasible estimate of $\boldsymbol{\beta}$. As for the nonparametric fit in (2.1), the local linear estimator given in (1.9) with a multiplicative triweight kernel was considered. The bandwidth was taken as a diagonal matrix $\mathbf{H} = \text{diag}(h_1, h_2)$, being the values of h_1 and h_2 different. The range of bandwidths was selected taking into account the MGCV bandwidth (Francisco-Fernandez and Opsomer, 2005; Francisco-Fernández et al., 2012), given in (1.33), $\mathbf{H} = \text{diag}(403.19, 226.20)$.

Figure 2.8 shows the smoothed version of the parametric (left panel) and the nonparametric (right panel) regression estimators using the corrected generalized cross-validation bandwidth for the level of piezometric-head in the area of study. These regression surfaces are compared in the proposed test statistic. Figure 2.9 shows the p-values of the test using the so-called significance trace (Bowman and Azzalini, 1997), that is, the proportions of empirical rejections for different bandwidths. Taking into account this plot, there are no evidences against a linear spatial regression. Note that smaller bandwidths than those considered should not be taken to avoid boundary problems.

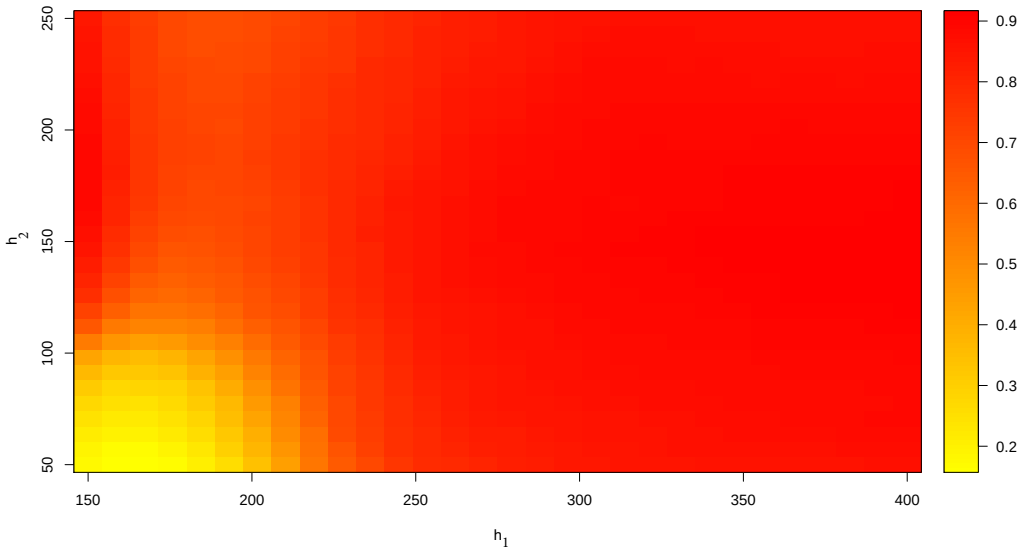

Figure 2.9: For the Wolfcamp aquifer dataset, p-values of the test for different values of h_1 and h_2, considering the parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}$ as the null hypothesis.

2.6 Proofs of the main results

In this section, proofs of Theorems 2.1 and 2.2 are provided. The proof of Theorem 2.1 is shown only for $T_{n,1}$, being similar for $T_{n,0}$. To derive the proof of Theorem 2.2, it is sufficient to follow similar arguments to those used in the proof of Theorem 2.1, but employing Riemann approximations of sums by integrals. For this reason, only a sketch of the proof of Theorem 2.2 with some highlights is provided. For sake of simplicity, the proof of this result is only presented for the case of $T_{n,0}$.

Proof of Theorem 2.1 Under assumptions (A1)–(A7), (H2), (K1) and (K2), the asymptotic distribution of the test statistic given in (2.1), comparing the nonparametric and the smooth parametric estimators, given in (1.9) and (2.2), respectively, using an L_2-distance, is derived. For simplicity, in this proof, $T_{n,1}$ will be denoted by T_n.

The test statistic given in (2.1) can be written as:

$$\begin{aligned} T_n &= n|\mathbf{H}|^{1/2} \int [\hat{m}_{\mathbf{H}}(\mathbf{x};1) - \hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};1)]^2 w(\mathbf{x}) d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \left\{ \mathbf{e}_1^{\mathrm{T}} \left(n^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}} \right)^{-1} \right. \\ &\quad \left. \cdot \frac{1}{n} \sum_{i=1}^{n} [1, (\mathbf{X}_i - \mathbf{x})^{\mathrm{T}}] K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) [Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \left[\mathbf{e}_1^{\mathrm{T}} \left(n^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}} \right)^{-1} \right. \\ &\quad \left. \cdot \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) [Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \\ \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x}) [Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \end{pmatrix} \right]^2 w(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

Using (A.1), (A.2) and (A.3) of Lemma A.1, and taking into account that for every $\eta > 0$, $\frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) = f(\mathbf{x}) + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta})$ uniformly in $\mathbf{x}$ (see Härdle and Mammen, 1993), it follows that

$$\begin{aligned} &(n^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1} \\ &= \begin{pmatrix} \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) & \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x})^{\mathrm{T}} \\ \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x}) & \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x})^{\mathrm{T}} \end{pmatrix}^{-1} \\ &= \begin{pmatrix} f(\mathbf{x}) + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta}) & \mu_2(K) \boldsymbol{\nabla}^{\mathrm{T}} f(\mathbf{x}) \mathbf{H}^2 + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{H}^2) \\ \mu_2(K) \mathbf{H}^2 \boldsymbol{\nabla} f(\mathbf{x}) + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{H}^2) & \mu_2(K) f(\mathbf{x}) \mathbf{H}^2 + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{H} \mathbf{1}_{d \times d} \mathbf{H}) \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{1}{f(\mathbf{x})} + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta}) & \frac{-\boldsymbol{\nabla}^{\mathrm{T}} f(\mathbf{x})}{f^2(\mathbf{x})} + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{1}_d^{\mathrm{T}}) \\ \frac{-\boldsymbol{\nabla} f(\mathbf{x})}{f^2(\mathbf{x})} + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{1}_d) & \frac{1}{\mu_2(K) f(\mathbf{x}) \mathbf{H}^2} + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta} \mathbf{H} \mathbf{1}_{d \times d} \mathbf{H}) \end{pmatrix}. \end{aligned}$$

The test statistic can be now written as:

$$\begin{aligned} T_n &= n|\mathbf{H}|^{1/2} \int \left\{ \frac{1}{n f(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) [Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right. \\ &\quad \left. - \boldsymbol{\nabla} f(\mathbf{x}) \frac{1}{n f^2(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) (\mathbf{X}_i - \mathbf{x}) [Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\ &\quad + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta}) \\ &= T_{n1} + T_{n2} + 2T_{n12} + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta}), \end{aligned} \tag{2.12}$$

with

$$T_{n1} = n|\mathbf{H}|^{1/2} \int \left\{ \frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})[Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x},$$

$$T_{n2} = n|\mathbf{H}|^{1/2} \int \left\{ \boldsymbol{\nabla} f(\mathbf{x}) \frac{1}{nf^2(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})[Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x},$$

and the T_{n12} term is the integral of the cross product.

Regarding T_{n1}, taking into account that the regression functions considered are of the form $m = m_{\boldsymbol{\beta}_0} + n^{-1/2}|\mathbf{H}|^{-1/4} g$ (see Section 2.2), one gets

$$\begin{aligned}
T_{n1} &= n|\mathbf{H}|^{1/2} \int \left\{ \frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})[Z_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \left\{ \frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})[m(\mathbf{X}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \left\{ \frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})[m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)] \right. \\
&\quad \left. + \frac{1}{nf(\mathbf{x})} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})[n^{-1/2}|\mathbf{H}|^{-1/4} g(\mathbf{X}_i) + \varepsilon_i] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} [I_1(\mathbf{x}) + I_2(\mathbf{x}) + I_3(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x},
\end{aligned}$$

where

$$\begin{aligned}
I_1(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) [m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)], \\
I_2(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) n^{-1/2}|\mathbf{H}|^{-1/4} g(\mathbf{X}_i), \\
I_3(\mathbf{x}) &= \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \varepsilon_i.
\end{aligned}$$

With respect to the term $I_1(\mathbf{x})$, using assumptions (A1), (A2), (A4), (A5) and (K1), and given that the difference $m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{\boldsymbol{\beta}_0}(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ uniformly in $\mathbf{x}$ (see

Section 2.2), it is obtained that

$$\begin{aligned} & n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} I_1^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}} \left(\mathbf{X}_i - \mathbf{x}\right) \left[m_{\boldsymbol{\beta}_0}(\mathbf{X}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) \right] \right\}^2 w(\mathbf{x}) d\mathbf{x} \\ &= \mathcal{O}_{\mathbb{P}}(|\mathbf{H}|^{1/2}). \end{aligned} \tag{2.13}$$

As for the term $I_2(\mathbf{x})$, taking into account (A.1) of Lemma A.1, it follows that

$$\begin{aligned} & n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} I_2^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \left[\frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}} \left(\mathbf{X}_i - \mathbf{x}\right) n^{-1/2} |\mathbf{H}|^{-1/4} g(\mathbf{X}_i) \right]^2 w(\mathbf{x}) d\mathbf{x} \\ &= \int \frac{1}{f^2(\mathbf{x})} \left\{ \int K\left(\mathbf{p}\right) g(\mathbf{x} + \mathbf{H}\mathbf{p}) f(\mathbf{x} + \mathbf{H}\mathbf{p}) d\mathbf{p} + o_{\mathbb{P}}(1) \right\}^2 w(\mathbf{x}) d\mathbf{x} \\ &= \int \frac{1}{f^2(\mathbf{x})} \left\{ \int K\left(\mathbf{p}\right) g(\mathbf{x} + \mathbf{H}\mathbf{p}) [f(\mathbf{x}) + o(1)] d\mathbf{p} \right\}^2 w(\mathbf{x}) d\mathbf{x} \cdot [1 + o_{\mathbb{P}}(1)] \\ &= \int \left[\int K_{\mathbf{H}} \left(\mathbf{u} - \mathbf{x}\right) g(\mathbf{u}) d\mathbf{u} \right]^2 w(\mathbf{x}) d\mathbf{x} \cdot [1 + o_{\mathbb{P}}(1)] \\ &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x} \cdot [1 + o_{\mathbb{P}}(1)]. \end{aligned} \tag{2.14}$$

The leading term of (2.14) is the term $b_{1\mathbf{H}}$ in Theorem 2.1. Finally, the term $I_3(\mathbf{x})$, associated with the error component of the model, can be split as:

$$\begin{aligned} & n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} I_3^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \left[\frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}} \left(\mathbf{X}_i - \mathbf{x}\right) \varepsilon_i \right]^2 w(\mathbf{x}) d\mathbf{x} \\ &= I_{31} + I_{32}, \end{aligned}$$

where

$$
\begin{aligned}
I_{31}(\mathbf{x}) &= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 \left(\mathbf{X}_i - \mathbf{x}\right) \varepsilon_i^2 w(\mathbf{x}) d\mathbf{x} \\
I_{32}(\mathbf{x}) &= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} K_{\mathbf{H}} \left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}} \left(\mathbf{X}_j - \mathbf{x}\right) \varepsilon_i \varepsilon_j w(\mathbf{x}) d\mathbf{x}.
\end{aligned}
$$

Close expressions of I_{31} and I_{32} can be obtained computing the expectation and the variance of these terms. For doing so, general results on the conditional expectation and conditional variance can be used. Specifically, given two random variables X and Y, it is known that $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$ and $\mathbb{V}\text{ar}(X) = \mathbb{E}[\mathbb{V}\text{ar}(X|Y)] + \mathbb{V}\text{ar}[\mathbb{E}(X|Y)]$.

For I_{31}, using the result for the conditional mean, it follows that $\mathbb{E}(I_{31}) = \mathbb{E}[\mathbb{E}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)]$. Firstly,

$$
\begin{aligned}
&\mathbb{E}(|\mathbf{H}|^{1/2} I_{31} | \mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&\quad = \mathbb{E}\left[n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 \left(\mathbf{X}_i - \mathbf{x}\right) \varepsilon_i^2 w(\mathbf{x}) d\mathbf{x} | \mathbf{X}_1, \ldots, \mathbf{X}_n \right] \\
&\quad = \sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 \left(\mathbf{X}_i - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x}. \qquad (2.15)
\end{aligned}
$$

Using the previous expression and (A.4) of Lemma A.1, one gets that,

$$
\begin{aligned}
\mathbb{E}(|\mathbf{H}|^{1/2} I_{31}) &= \mathbb{E}[\mathbb{E}(|\mathbf{H}|^{1/2} I_{31} | \mathbf{X}_1, \ldots, \mathbf{X}_n)] \\
&= \mathbb{E}\left[\sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 \left(\mathbf{X}_i - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x} \right] \\
&= \sigma^2 |\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} |\mathbf{H}|^{-1} K^{[2]}(\mathbf{0}) [f(\mathbf{x}) + o(1)] w(\mathbf{x}) d\mathbf{x} \\
&= \sigma^2 K^{[2]}(\mathbf{0}) \int \frac{w(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot [1 + o(1)], \qquad (2.16)
\end{aligned}
$$

since $R(K) = \int K^2(\mathbf{u}) d\mathbf{u} = K^{[2]}(\mathbf{0})$.

On the other hand, $\mathbb{V}\text{ar}(I_{31}) = \mathbb{E}[\mathbb{V}\text{ar}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)] + \mathbb{V}\text{ar}[\mathbb{E}(I_{31}|\mathbf{X}_1, \ldots, \mathbf{X}_n)]$.

Using assumption (A6), it is obtained that

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}(|\mathbf{H}|^{1/2} I_{31} | \mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&= \mathbb{V}\mathrm{ar}\left[n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i=1}^{n} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x}) \, \varepsilon_i^2 w(\mathbf{x}) d\mathbf{x} | \mathbf{X}_1, \ldots, \mathbf{X}_n \right] \\
&= \frac{|\mathbf{H}|^2}{n^2} \sum_{i,j=1}^{n} \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x}) \, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t}) \, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \mathbb{C}\mathrm{ov}(\varepsilon_i^2, \varepsilon_j^2) \\
&= \frac{2|\mathbf{H}|^2}{n^2} \sum_{i,j=1}^{n} \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x}) \, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t}) \, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} [\mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j)]^2 \\
&= \frac{2\sigma^4}{n^2} |\mathbf{H}|^2 \sum_{i,j=1}^{n} \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x}) \, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t}) \, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&\quad \cdot \rho_n^2 (\mathbf{X}_i - \mathbf{X}_j).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}[\mathbb{V}\mathrm{ar}(|\mathbf{H}|^{1/2} I_{31} | \mathbf{X}_1, \ldots, \mathbf{X}_n)] \\
&= \mathbb{E}\left[\frac{2\sigma^4}{n^2} |\mathbf{H}|^2 \sum_{i,j=1}^{n} \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} K_{\mathbf{H}}^2 (\mathbf{X}_i - \mathbf{x}) \, K_{\mathbf{H}}^2 (\mathbf{X}_j - \mathbf{t}) \, w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \right] \\
&= 2\sigma^4 |\mathbf{H}|^2 \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} \iint K_{\mathbf{H}}^2 (\mathbf{u} - \mathbf{x}) \, K_{\mathbf{H}}^2 (\mathbf{v} - \mathbf{t}) \, \rho_n^2 (\mathbf{u} - \mathbf{v}) f(\mathbf{u}) f(\mathbf{v}) d\mathbf{u} d\mathbf{v} \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 2\sigma^4 \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} \iint K^2(\mathbf{p}) K^2(\mathbf{q}) \rho_n^2 [\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] f(\mathbf{x} + \mathbf{H}\mathbf{p}) \\
&\quad \cdot f(\mathbf{t} + \mathbf{H}\mathbf{q}) d\mathbf{p} d\mathbf{q} w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \cdot [1 + o(1)] \\
&= 2\sigma^4 \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} \iint K^2(\mathbf{p}) K^2(\mathbf{q}) \rho_n^2 [\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] d\mathbf{p} d\mathbf{q} \\
&\quad \cdot f(\mathbf{x}) f(\mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \cdot [1 + o(1)] \\
&= 2\sigma^4 \iiiint \frac{K^2(\mathbf{p}) K^2(\mathbf{q})}{f(\mathbf{x}) f(\mathbf{t})} w(\mathbf{x}) w(\mathbf{t}) \rho_n^2 [\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] d\mathbf{p} d\mathbf{q} d\mathbf{x} d\mathbf{t} \cdot [1 + o(1)] \\
&= 2\sigma^4 |\mathbf{H}| \iiiint \frac{K^2(\mathbf{p}) K^2(\mathbf{q})}{f(\mathbf{x}) f(\mathbf{x} + \mathbf{H}\mathbf{u})} w(\mathbf{x}) w(\mathbf{x} + \mathbf{H}\mathbf{u}) \rho_n^2 [\mathbf{H}(\mathbf{p} - \mathbf{q} - \mathbf{u})] d\mathbf{p} d\mathbf{q} d\mathbf{x} d\mathbf{u} \\
&\quad \cdot [1 + o(1)] \\
&= 2\sigma^4 |\mathbf{H}| \iiiint \frac{K^2(\mathbf{p}) K^2(\mathbf{q})}{f^2(\mathbf{x})} w^2(\mathbf{x}) \rho_n^2 [\mathbf{H}(\mathbf{p} - \mathbf{q} - \mathbf{u})] d\mathbf{p} d\mathbf{q} d\mathbf{x} d\mathbf{u} \cdot [1 + o(1)].
\end{aligned}
$$

Let

$$j_n(\mathbf{q},\mathbf{u}) = n|\mathbf{H}| \int K^2(\mathbf{p})\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q}-\mathbf{u})]d\mathbf{p}.$$

Notice that,

$$|j_n(\mathbf{q},\mathbf{u})| \;\leq\; K_M^2\{n|\mathbf{H}| \int_{\|\mathbf{p}\|\leq 1} |\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q}-\mathbf{u})]|d\mathbf{p}\} \leq K_M^2[n \int |\rho_n(\mathbf{t})|d\mathbf{t}],$$

where $K_M = \max_{\mathbf{x}}[K(\mathbf{x})]$. Using assumption (A3), it follows that

$$|j_n(\mathbf{q},\mathbf{u})| \leq K_M^2 \rho_M,$$

where $\rho_M = \max_{\mathbf{x}}[\rho(\mathbf{x})]$, and therefore,

$$\mathbb{E}[\mathbb{V}\mathrm{ar}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] \leq 2n^{-1}\sigma^4 K_M^2 \rho_M \iiint \frac{K^2(\mathbf{q})}{f^2(\mathbf{x})} w^2(\mathbf{x}) d\mathbf{q}d\mathbf{x}d\mathbf{u} \cdot [1+o(1)].$$

Using assumptions (A4), (A5) and (K1), one gets that

$$\mathbb{E}[\mathbb{V}\mathrm{ar}(I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] = o_{\mathbb{P}}(1). \tag{2.17}$$

On the other hand, using (2.15), (A.4) of Lemma A.1 and (H2), it follows that

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}[\mathbb{E}(|\mathbf{H}|^{1/2} I_{31}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] \\
&= \mathbb{V}\mathrm{ar}\left[\sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i=1}^n K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})\, w(\mathbf{x})d\mathbf{x}\right] \\
&= \sum_{i=1}^n \mathbb{V}\mathrm{ar}\left[\sigma^2 n|\mathbf{H}| \int \frac{1}{f^2(\mathbf{x})}\frac{1}{n^2} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})\, w(\mathbf{x})d\mathbf{x}\right] \\
&\leq \sigma^4|\mathbf{H}|^2 \sum_{i=1}^n \mathbb{E}\left[\iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\frac{1}{n^2} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})\, K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{t})\, w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\right] \\
&= o_{\mathbb{P}}(1).
\end{aligned} \tag{2.18}$$

Now, considering (2.17) and (2.18), it is obtained that

$$\mathbb{V}\mathrm{ar}(|\mathbf{H}|^{1/2} I_{31}) = o_{\mathbb{P}}(1). \tag{2.19}$$

Using (2.16) and (2.19), it follows that

$$I_{31} = |\mathbf{H}|^{-1/2}\sigma^2 K^{[2]}(\mathbf{0}) \int \frac{w(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot [1 + o_{\mathbb{P}}(1)]. \tag{2.20}$$

Taking into account assumption (H2), the leading term of (2.20) corresponds to the first term of $b_{0\mathbf{H}}$ in Theorem 2.1. Now, consider the term

$$I_{32} \;=\; n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) \varepsilon_i \varepsilon_j w(\mathbf{x}) d\mathbf{x}.$$

Let

$$\kappa_{ij} \;=\; n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) \varepsilon_i \varepsilon_j w(\mathbf{x}) d\mathbf{x},$$

thus,

$$I_{32} = \sum_{i \neq j} \kappa_{ij},$$

and this can be seen as a U-statistic with degenerate kernel.

To obtain the asymptotic normality of I_{32}, Theorem 2 of Kim et al. (2013) will be applied. The central limit theorem for degenerate reduced U-statistics under α-mixing was derived in that work. The assumptions of this result hold (specifically, assumption (A7)) and the expectation and the variance of I_{32} should be computed.

Proceeding as for I_{31}, it follows that $\mathbb{E}(I_{32}) = \mathbb{E}[\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n)]$. First,

$$\begin{aligned}
&\mathbb{E}(I_{32}|\mathbf{X}_1, \ldots, \mathbf{X}_n) \\
&= \mathbb{E}\left[n|\mathbf{H}|^{1/2} \frac{1}{n^2} \sum_{i \neq j} \int \frac{1}{f^2(\mathbf{x})} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x} \cdot \varepsilon_i \varepsilon_j | \mathbf{X}_1, \ldots, \mathbf{X}_n \right] \\
&= n|\mathbf{H}|^{1/2} \frac{1}{n^2} \int \frac{1}{f^2(\mathbf{x})} \sum_{i \neq j} \mathbb{E}(\varepsilon_i \varepsilon_j) K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2} \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n^2} \sum_{i \neq j} \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j) K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x} \\
&= |\mathbf{H}|^{1/2} \sigma^2 \int \frac{1}{f^2(\mathbf{x})} \frac{1}{n} \sum_{i \neq j} \rho_n(\mathbf{X}_i - \mathbf{X}_j) K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) w(\mathbf{x}) d\mathbf{x}.
\end{aligned} \tag{2.21}$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}[\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] \\
&= \mathbb{E}\left[n|\mathbf{H}|^{1/2}\sigma^2\int\frac{1}{f^2(\mathbf{x})}\frac{1}{n^2}\sum_{i\neq j}\rho_n(\mathbf{X}_i-\mathbf{X}_j)K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)w(\mathbf{x})d\mathbf{x}\right] \\
&= n|\mathbf{H}|^{1/2}\sigma^2\int\frac{1}{f^2(\mathbf{x})}\left\{\frac{n-1}{n}\iint K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{t}\right)\rho_n(\mathbf{u}-\mathbf{v})f(\mathbf{u})f(\mathbf{v})d\mathbf{u}d\mathbf{v}\right\} \\
&\quad \cdot w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2}\sigma^2\int\frac{1}{f^2(\mathbf{x})}\left\{\frac{n-1}{n}\iint K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]\right. \\
&\quad \left.\cdot f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{q})d\mathbf{p}d\mathbf{q}\right\}w(\mathbf{x})d\mathbf{x} \\
&= n|\mathbf{H}|^{1/2}\sigma^2\int\frac{1}{f^2(\mathbf{x})}\left\{\frac{n-1}{n}f^2(\mathbf{x})\iint K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}+o(1)\right\} \\
&\quad \cdot w(\mathbf{x})d\mathbf{x} \\
&= \frac{n-1}{n}|\mathbf{H}|^{-1/2}\sigma^2\int\left\{n|\mathbf{H}|\iint K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}+o(1)\right\} \\
&\quad \cdot w(\mathbf{x})d\mathbf{x}.
\end{aligned}
$$

Using Proposition A.2, it follows that

$$\mathbb{E}\left(I_{32}\right)=|\mathbf{H}|^{-1/2}\sigma^2K^{[2]}(\mathbf{0})\rho_c\int w(\mathbf{x})d\mathbf{x}\cdot[1+o(1)], \tag{2.22}$$

corresponding to the second term of $b_{0\mathbf{H}}$ in Theorem 2.1.

The variance of I_{32} can be computed considering that

$$\mathbb{V}\mathrm{ar}(I_{32}) = \mathbb{E}[\mathbb{V}\mathrm{ar}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)]+\mathbb{V}\mathrm{ar}[\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)]. \tag{2.23}$$

Let

$$W_{ij} = \int\frac{1}{f^2(\mathbf{x})}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)w(\mathbf{x})d\mathbf{x},$$

thus,

$$\begin{aligned}\mathbb{V}\text{ar}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n) &= \mathbb{V}\text{ar}\left(n^{-1}|\mathbf{H}|^{1/2}\sum_{i\neq j}W_{ij}\varepsilon_i\varepsilon_j|\mathbf{X}_1,\ldots,\mathbf{X}_n\right)\\ &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}W_{ij}W_{kl}\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j,\varepsilon_k\varepsilon_l)\\ &= T_{31}+T_{32}+T_{33}, \end{aligned} \tag{2.24}$$

where

$$\begin{aligned}T_{31} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}^2\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_j),\\ T_{32} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=i+2}^{n}W_{ij}W_{il}\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_l),\\ T_{33} &= 4n^{-2}|\mathbf{H}|\sum_{\text{all different indices } i,\ j,\ k,\ l}W_{ij}W_{kl}\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j,\varepsilon_k\varepsilon_l).\end{aligned}$$

First, when $i=k$ and $j=l$, the total number of terms is $n(n-1)/2$. Second, when one of the i and j is equal to one of the k and l (without loss of generality, assume $i=k$ and $j\neq l$), the total number of terms can be bounded by n^3. Finally, when i,j,k, and l are all different, the total number of terms can be bounded by n^4.

The expected value of $\mathbb{V}\text{ar}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ given in (2.24) is computed by calculating the mean of the terms T_{31}, T_{32}, and T_{33},

$$\mathbb{E}[\mathbb{V}\text{ar}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] = \mathbb{E}(T_{31})+\mathbb{E}(T_{32})+\mathbb{E}(T_{33}). \tag{2.25}$$

As for T_{31}, using assumption (A6), this term can be split as:

$$\begin{aligned}T_{31} &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}^2\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j,\varepsilon_i\varepsilon_j)\\ &= 4n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}W_{ij}^2[\sigma^4+\mathbb{C}\text{ov}^2(\varepsilon_i,\varepsilon_j)]\\ &= T_{311}+T_{312},\end{aligned}$$

where

$$T_{311} = 4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} w(\mathbf{x})w(\mathbf{t}) \cdot \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) d\mathbf{x} d\mathbf{t},$$

and

$$T_{312} = 4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} w(\mathbf{x})w(\mathbf{t}) \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \cdot K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) \rho_n^2(\mathbf{X}_i - \mathbf{X}_j) d\mathbf{x} d\mathbf{t}.$$

First,

$$\begin{aligned}
\mathbb{E}(T_{311}) &= \mathbb{E}\Bigg[4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} w(\mathbf{x})w(\mathbf{t}) \\
&\quad \cdot \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{t}) d\mathbf{x} d\mathbf{t}\Bigg] \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \left[\int K_{\mathbf{H}}(\mathbf{u} - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{u} - \mathbf{t})\, f(\mathbf{u}) d\mathbf{u}\right]^2 \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \\
&\quad \cdot \left[|\mathbf{H}|^{-1} \int K(\mathbf{p})\, K[\mathbf{p} - \mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})] f(\mathbf{t} + \mathbf{H}\mathbf{p}) d\mathbf{p}\right]^2 w(\mathbf{x})w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \left[|\mathbf{H}|^{-1} K^{[2]}[\mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})][f(\mathbf{t}) + o(1)]\right]^2 \\
&\quad \cdot w(\mathbf{x})w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 2\sigma^4|\mathbf{H}|\frac{n-1}{n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-2}\{K^{[2]}[\mathbf{H}^{-1}(\mathbf{x} - \mathbf{t})]\}^2 f^2(\mathbf{t}) w(\mathbf{x})w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&\quad \cdot [1 + o(1)] \\
&= 2\sigma^4 \frac{n-1}{n} \iint \frac{1}{f^2(\mathbf{x})} [K^{[2]}(\mathbf{p})]^2 w(\mathbf{x})w(\mathbf{x} + \mathbf{H}\mathbf{p}) d\mathbf{x} d\mathbf{p} \cdot [1 + o(1)] \\
&= 2\sigma^4 K^{[4]}(\mathbf{0}) \int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})} d\mathbf{x} \cdot [1 + o(1)]. \qquad (2.26)
\end{aligned}$$

Similarly for T_{312}, using assumptions (A4), (A5) and (K1), this term becomes

$$\begin{aligned}
\mathbb{E}(T_{312}) &= \mathbb{E}\Big[4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} w(\mathbf{x})w(\mathbf{t}) \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right) \\
&\quad \cdot K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{t}) \rho_n^2(\mathbf{X}_i-\mathbf{X}_j) d\mathbf{x} d\mathbf{t}\Big] \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \iint K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{t}\right) K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{x}\right) \\
&\quad \cdot K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{t}\right) \rho_n^2(\mathbf{u}-\mathbf{v}) f(\mathbf{u}) f(\mathbf{v}) d\mathbf{u} d\mathbf{v} w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-2} \iint K[-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] \\
&\quad \cdot K[-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] K\left(\mathbf{p}\right) K\left(\mathbf{q}\right) f(\mathbf{t}+\mathbf{H}\mathbf{p}) f(\mathbf{t}+\mathbf{H}\mathbf{q}) \rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p} d\mathbf{q} \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4|\mathbf{H}|\frac{n-1}{2n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} |\mathbf{H}|^{-2} f^2(\mathbf{t}) \iint K[-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] \\
&\quad \cdot K[-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] K(\mathbf{p}) K(\mathbf{q}) \rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p} d\mathbf{q} \cdot [1+o(1)] \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 2\frac{n-1}{n}\sigma^4 \iiiint \frac{1}{f^2(\mathbf{x})} K\left(\mathbf{p}+\mathbf{u}\right) K\left(\mathbf{q}+\mathbf{u}\right) K\left(\mathbf{p}\right) K\left(\mathbf{q}\right) \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{x}+\mathbf{H}\mathbf{u}) \rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p} d\mathbf{q} d\mathbf{x} d\mathbf{u} \cdot [1+o(1)] \\
&\leq 2\frac{n-1}{n^2|\mathbf{H}|}\frac{\sigma^4 K_M^4 w_M^2}{f_M^2} \int_{\|\mathbf{q}\|\leq 1} \{n|\mathbf{H}| \int_{\|\mathbf{p}\|\leq 1} \rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p}\} d\mathbf{q} \cdot [1+o(1)],
\end{aligned}$$

where f_M and w_M denote the upper bounds of f and w, respectively (assumption (A4)). Since

$$n|\mathbf{H}| \int_{\|\mathbf{p}\|\leq 1} \rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p} \leq n \int |\rho_n(\mathbf{t})| d\mathbf{t} \leq C_1,$$

it is obtained that

$$\begin{aligned}
\mathbb{E}(T_{312}) &\leq 2\frac{\sigma^4 K_M^4 w_M^2}{f_M^2} \frac{C_1}{n|\mathbf{H}|} \frac{n-1}{n} \times \text{Volume of the ball } \{\mathbf{u} : \|\mathbf{u}\| \leq 1\} \\
&= \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}). \qquad (2.27)
\end{aligned}$$

Then, from (2.26) and (2.27), it follows that

$$\mathbb{E}(T_{31}) = 2\sigma^4 K^{[4]}(\mathbf{0}) \int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})} d\mathbf{x} \cdot [1 + o(1)] + \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}). \tag{2.28}$$

With respect to the term T_{32} (corresponding to the case with $i = k$ and $j \neq l$ in (2.24)), using assumption (A6), it follows that

$$\begin{aligned} T_{32} &= 4n^{-2}|\mathbf{H}| \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{j=i+2}^{n} W_{ij} W_{il} \mathbb{C}\mathrm{ov}(\varepsilon_i \varepsilon_j, \varepsilon_i \varepsilon_l) \\ &= 4n^{-2}|\mathbf{H}| \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{j=i+2}^{n} W_{ij} W_{il} [\mathbb{V}\mathrm{ar}(\varepsilon_i) \mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_l) + \mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_l) \mathbb{C}\mathrm{ov}(\varepsilon_j, \varepsilon_i)] \\ &= T_{321} + T_{322}, \end{aligned}$$

where

$$\begin{aligned} T_{321} &= 4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} w(\mathbf{x}) w(\mathbf{t}) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{l=i+2}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) \rho_n(\mathbf{X}_j - \mathbf{X}_l) d\mathbf{x} d\mathbf{t}, \\ T_{322} &= 4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} w(\mathbf{x}) w(\mathbf{t}) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{l=i+2}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) \rho_n(\mathbf{X}_i - \mathbf{X}_l) \rho_n(\mathbf{X}_j - \mathbf{X}_i) d\mathbf{x} d\mathbf{t}. \end{aligned}$$

Firstly,

$$\begin{aligned} \mathbb{E}(T_{321}) &= \mathbb{E}\Bigg[4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} w(\mathbf{x}) w(\mathbf{t}) \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{l=i+2}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) \rho_n(\mathbf{X}_j - \mathbf{X}_l) d\mathbf{x} d\mathbf{t} \Bigg] \\ &= 4\sigma^4 \frac{(n^2 - 3n + 2)}{n} |\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x}) f^2(\mathbf{t})} \int K_{\mathbf{H}}(\mathbf{w} - \mathbf{x}) K_{\mathbf{H}}(\mathbf{w} - \mathbf{t}) f(\mathbf{w}) d\mathbf{w} \\ &\quad \cdot \left\{ \iint K_{\mathbf{H}}(\mathbf{u} - \mathbf{x}) K_{\mathbf{H}}(\mathbf{v} - \mathbf{t}) \rho_n(\mathbf{u} - \mathbf{v}) f(\mathbf{u}) f(\mathbf{v}) d\mathbf{u} d\mathbf{v} \right\} \\ &\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t}. \end{aligned}$$

Using the assumption (A3) and Proposition A.2, it follows that

$$\begin{aligned}
\mathbb{E}(T_{321}) &= 4\sigma^4\frac{(n^2-3n+2)}{n}\iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\int K(\mathbf{r})\,K[\mathbf{r}-\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]f(\mathbf{t}+\mathbf{H}\mathbf{r})d\mathbf{r}\\
&\quad\cdot\left\{\iint K(\mathbf{p})\,K(\mathbf{q})\,\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q})]f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{t}+\mathbf{H}\mathbf{q})d\mathbf{p}d\mathbf{q}\right\}\\
&\quad\cdot w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= 4\sigma^4\frac{(n^2-3n+2)}{n}\iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K^{[2]}[\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]f(\mathbf{t})\cdot[1+o(1)]\\
&\quad\cdot\left\{\iint K(\mathbf{p})\,K(\mathbf{q})\,\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}f(\mathbf{x})f(\mathbf{t})\cdot[1+o(1)]\right\}\\
&\quad\cdot w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= 4\sigma^4\frac{(n^2-3n+2)}{n}\iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K^{[2]}[\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]f(\mathbf{t})\\
&\quad\cdot\iint K(\mathbf{p})\,K(\mathbf{q})\,\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}f(\mathbf{x})f(\mathbf{t})w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&\quad\cdot[1+o(1)]\\
&= 4\sigma^4\frac{(n^2-3n+2)}{n}\iint\frac{1}{f(\mathbf{x})}K^{[2]}[\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]\\
&\quad\cdot\iint K(\mathbf{p})\,K(\mathbf{q})\,\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\cdot[1+o(1)]\\
&= 4\sigma^4\frac{(n^2-3n+2)}{n}|\mathbf{H}|\iiint\frac{1}{f(\mathbf{x})}K^{[2]}(\mathbf{r})\,K(\mathbf{p})\,K(\mathbf{q})\,w(\mathbf{x})w(\mathbf{x}-\mathbf{H}\mathbf{r})\\
&\quad\cdot\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{r})]d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{r}\cdot[1+o(1)]\\
&= 4\sigma^4\frac{(n^2-3n+2)}{n^2}\iiint\frac{1}{f(\mathbf{x})}K^{[2]}(\mathbf{r})\,K(\mathbf{q})\,w^2(\mathbf{x})\\
&\quad\cdot\{n|\mathbf{H}|\int K(\mathbf{p})\,\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q}+\mathbf{r})]d\mathbf{p}\}d\mathbf{q}d\mathbf{x}d\mathbf{r}\cdot[1+o(1)]\\
&= 4\sigma^4\rho_c\iiint\frac{1}{f(\mathbf{x})}K^{[2]}(\mathbf{r})\,K(\mathbf{q})\,K(\mathbf{r}-\mathbf{q})w^2(\mathbf{x})d\mathbf{q}d\mathbf{x}d\mathbf{r}\cdot[1+o(1)]\\
&= 4\sigma^4\rho_c\iint\frac{1}{f(\mathbf{x})}[K^{[2]}(\mathbf{r})]^2w^2(\mathbf{x})d\mathbf{r}d\mathbf{x}\cdot[1+o(1)]\\
&= 4\sigma^4K^{[4]}(\mathbf{0})\rho_c\int\frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}\cdot[1+o(1)]. \qquad (2.29)
\end{aligned}$$

Similarly, taking into account that K is bounded and assumption (A3), the ex-

pected value of T_{322} becomes

$$\begin{aligned}
\mathbb{E}(T_{322}) &= \mathbb{E}\Big[4\sigma^4 n^{-2}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} w(\mathbf{x})w(\mathbf{t}) \sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{l=i+2}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right) \\
&\quad \cdot K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right) K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l-\mathbf{t}) \rho_n(\mathbf{X}_i-\mathbf{X}_l)\rho_n(\mathbf{X}_j-\mathbf{X}_i) d\mathbf{x}d\mathbf{t}\Big] \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n}|\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \iiint K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{t}\right) \\
&\quad \cdot K_{\mathbf{H}}\left(\mathbf{y}-\mathbf{t}\right) \rho_n(\mathbf{u}-\mathbf{y})\rho_n(\mathbf{u}-\mathbf{v}) f(\mathbf{u})f(\mathbf{v})f(\mathbf{y}) d\mathbf{u}d\mathbf{v}d\mathbf{y} w(\mathbf{x})w(\mathbf{t}) d\mathbf{x}d\mathbf{t} \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \iiint K(\mathbf{p})K\left(\mathbf{q}\right) \\
&\quad \cdot K[\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] K\left(\mathbf{r}\right) f(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{q}) f(\mathbf{t}+\mathbf{H}\mathbf{r}) \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] \\
&\quad \cdot \rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})] d\mathbf{p}d\mathbf{q}d\mathbf{r} w(\mathbf{x})w(\mathbf{t}) d\mathbf{x}d\mathbf{t} \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} f^2(\mathbf{x})f(\mathbf{t}) \iiint K(\mathbf{p})K\left(\mathbf{q}\right) \\
&\cdot \quad K[\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] K\left(\mathbf{r}\right) \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] \rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})] d\mathbf{p}d\mathbf{q}d\mathbf{r} \\
&\quad \cdot [1+o(1)] w(\mathbf{x})w(\mathbf{t}) d\mathbf{x}d\mathbf{t} \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n} \int\cdots\int \frac{1}{f(\mathbf{t})} K\left(\mathbf{p}\right) K[\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})] K\left(\mathbf{q}\right) K\left(\mathbf{r}\right) \\
&\quad \cdot w(\mathbf{x})w(\mathbf{t}) \rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})] \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p}d\mathbf{q}d\mathbf{r}d\mathbf{x}d\mathbf{t} \cdot [1+o(1)] \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n}|\mathbf{H}| \int\cdots\int \frac{1}{f(\mathbf{t})} K\left(\mathbf{p}\right) K\left(\mathbf{p}+\mathbf{u}\right) K\left(\mathbf{q}\right) K\left(\mathbf{r}\right) \\
&\quad \cdot w(\mathbf{t}+\mathbf{H}\mathbf{u}) w(\mathbf{t}) \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})] \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{p}d\mathbf{q}d\mathbf{r}d\mathbf{u}d\mathbf{t} \cdot [1+o(1)] \\
&= 4\sigma^4 \frac{(n^2-3n+2)}{n^3}|\mathbf{H}|^{-1} \iiint \frac{1}{f(\mathbf{t})} K\left(\mathbf{p}\right) K\left(\mathbf{p}+\mathbf{u}\right) w^2(\mathbf{t}) \\
&\quad \cdot \{n|\mathbf{H}| \int K\left(\mathbf{r}\right) \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})] d\mathbf{r}\} \\
&\quad \cdot \{n|\mathbf{H}| \int K\left(\mathbf{q}\right) \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{q}\} d\mathbf{p}d\mathbf{u}d\mathbf{t} \cdot [1+o(1)].
\end{aligned}$$

Since

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{r})\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})] d\mathbf{r} = K(\mathbf{p}+\mathbf{u})\rho_c,$$

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{q})\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] d\mathbf{q} = K(\mathbf{p})\rho_c,$$

and taking into account that the functions K and w are bounded, and f is bounded away from zero, it follows that

$$\mathbb{E}(T_{322}) = \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}). \tag{2.30}$$

Then, from (2.29) and (2.30), one gets that

$$\mathbb{E}(T_{32}) = 4\sigma^4 K^{[4]}(\mathbf{0})\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})} d\mathbf{x} \cdot [1 + o(1)] + \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}). \tag{2.31}$$

Regarding the term T_{33} (when all i, j, k, l are different in (2.24)), using assumption (A6), it follows that

$$\begin{aligned} T_{33} &= 4n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} W_{ij}W_{kl}\mathbb{C}\text{ov}(\varepsilon_i\varepsilon_j, \varepsilon_k\varepsilon_l) \\ &= 4n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} W_{ij}W_{kl}[\mathbb{C}\text{ov}(\varepsilon_i, \varepsilon_k)\mathbb{C}\text{ov}(\varepsilon_j, \varepsilon_l) \\ &\quad +\mathbb{C}\text{ov}(\varepsilon_i, \varepsilon_l)\mathbb{C}\text{ov}(\varepsilon_j, \varepsilon_k)] \\ &= T_{331} + T_{332}, \end{aligned}$$

where

$$\begin{aligned} T_{331} &= 4\sigma^4 n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \rho_n(\mathbf{X}_i - \mathbf{X}_k) \rho_n(\mathbf{X}_j - \mathbf{X}_l), \end{aligned}$$

and

$$\begin{aligned} T_{332} &= 4\sigma^4 n^{-2}|\mathbf{H}| \sum_{\text{all different indices } i,\, j,\, k,\, l} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) \\ &\quad \cdot K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \rho_n(\mathbf{X}_i - \mathbf{X}_l) \rho_n(\mathbf{X}_j - \mathbf{X}_k). \end{aligned}$$

Using the assumption (A3), it follows that

$$\begin{aligned}
\mathbb{E}(T_{331}) &= \mathbb{E}\Bigg[4\sigma^4 n^{-2}|\mathbf{H}| \sum_{\text{all different } i,\, j,\, k,\, l} \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} K_{\mathbf{H}}\left(\mathbf{X}_i - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{X}_j - \mathbf{x}\right) \\
&\quad \cdot K_{\mathbf{H}}(\mathbf{X}_k - \mathbf{t}) K_{\mathbf{H}}(\mathbf{X}_l - \mathbf{t}) w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \rho_n(\mathbf{X}_i - \mathbf{X}_k) \rho_n(\mathbf{X}_j - \mathbf{X}_l)\Bigg] \\
&= 4\sigma^4 n^2 |\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \\
&\quad \cdot \left\{ \iint K_{\mathbf{H}}\left(\mathbf{u} - \mathbf{x}\right) K_{\mathbf{H}}\left(\mathbf{v} - \mathbf{t}\right) \rho_n(\mathbf{u} - \mathbf{v}) f(\mathbf{u}) f(\mathbf{v}) d\mathbf{u} d\mathbf{v} \right\}^2 \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4 n^2 |\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \\
&\quad \cdot \left\{ \iint K\left(\mathbf{p}\right) K\left(\mathbf{q}\right) \rho_n[\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] f(\mathbf{x} + \mathbf{H}\mathbf{p}) f(\mathbf{t} + \mathbf{H}\mathbf{q}) d\mathbf{p} d\mathbf{q} \right\}^2 \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4 n^2 |\mathbf{H}| \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})} \\
&\quad \cdot \left\{ f(\mathbf{x}) f(\mathbf{t}) \iint K\left(\mathbf{p}\right) K\left(\mathbf{q}\right) \rho_n[\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] d\mathbf{p} d\mathbf{q} \cdot [1 + o(1)] \right\}^2 \\
&\quad \cdot w(\mathbf{x}) w(\mathbf{t}) d\mathbf{x} d\mathbf{t} \\
&= 4\sigma^4 n^2 |\mathbf{H}| \int \cdots \int K\left(\mathbf{p}\right) K\left(\mathbf{q}\right) K\left(\mathbf{m}\right) K\left(\mathbf{r}\right) w(\mathbf{x}) w(\mathbf{t}) \\
&\cdot \quad \rho_n[\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{p} - \mathbf{q})] \rho_n[\mathbf{x} - \mathbf{t} + \mathbf{H}(\mathbf{m} - \mathbf{r})] d\mathbf{p} d\mathbf{q} d\mathbf{m} d\mathbf{r} d\mathbf{x} d\mathbf{t} \cdot [1 + o(1)] \\
&= 4\sigma^4 \iiiint K\left(\mathbf{q}\right) K\left(\mathbf{r}\right) w^2(\mathbf{x}) \{n|\mathbf{H}| \int K\left(\mathbf{m}\right) \rho_n[\mathbf{H}(\mathbf{m} - \mathbf{r} + \mathbf{u})] d\mathbf{m}\} \\
&\cdot \quad \{n|\mathbf{H}| \int K\left(\mathbf{p}\right) \rho_n[\mathbf{H}(\mathbf{p} - \mathbf{q} + \mathbf{u})] d\mathbf{p}\} d\mathbf{q} d\mathbf{r} d\mathbf{x} d\mathbf{u} \cdot [1 + o(1)].
\end{aligned}$$

Since

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{p}) \rho_n[\mathbf{H}(\mathbf{p} - \mathbf{q} + \mathbf{u})] d\mathbf{p} = K(\mathbf{q} - \mathbf{u}) \rho_c,$$

and

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{m}) \rho_n[\mathbf{H}(\mathbf{m} - \mathbf{r} + \mathbf{u})] d\mathbf{m} = K(\mathbf{r} - \mathbf{u}) \rho_c,$$

it follows that

$$\begin{aligned}\mathbb{E}(T_{331}) &= 4\sigma^4\rho_c^2 \iiiint K(\mathbf{q})\,K(\mathbf{u}-\mathbf{q})\,K(\mathbf{r})\,K(\mathbf{u}-\mathbf{r})\,w^2(\mathbf{x})d\mathbf{q}d\mathbf{r}d\mathbf{x}d\mathbf{u}\cdot[1+o(1)] \\ &= 4\sigma^4\rho_c^2 \iint [K^{[2]}(\mathbf{u})]^2 w^2(\mathbf{x})d\mathbf{x}d\mathbf{u}\cdot[1+o(1)] \\ &= 4\sigma^4\rho_c^2 K^{[4]}(\mathbf{0}) \int w^2(\mathbf{x})d\mathbf{x}\cdot[1+o(1)]. \end{aligned} \tag{2.32}$$

For symmetry, one gets that $\mathbb{E}(T_{332}) = \mathbb{E}(T_{331})$ and, therefore, using (2.32), it follows that

$$\mathbb{E}(T_{33}) = 8\sigma^4 K^{[4]}(\mathbf{0})\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\cdot[1+o(1)]. \tag{2.33}$$

So, from (2.25), (2.28), (2.31) and (2.33), it is obtained that

$$\begin{aligned}\mathbb{E}[\mathbb{V}\mathrm{ar}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] &= 2\sigma^4 K^{[4]}(\mathbf{0}) \int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x}\cdot[1+o(1)] + \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}) \\ &\quad +4\sigma^4 K^{[4]}(\mathbf{0})\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x}\cdot[1+o(1)] + \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}) \\ &\quad +8\sigma^4 K^{[4]}(\mathbf{0})\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\cdot[1+o(1)]. \end{aligned} \tag{2.34}$$

With respect to the $\mathbb{V}\mathrm{ar}[\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)]$, which is the second term in equation (2.23), denoting by:

$$\phi_{ij} = \int \frac{1}{f^2(\mathbf{x})} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\rho_n(\mathbf{X}_i-\mathbf{X}_j)w(\mathbf{x})d\mathbf{x},$$

and using the expression of the $\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)$ given in (2.21), it follows that

$$\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n) = |\mathbf{H}|^{1/2}\sigma^2\frac{1}{n}\sum_{i\neq j}\phi_{ij},$$

and, consequently,

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] &= \mathbb{V}\mathrm{ar}\left(|\mathbf{H}|^{1/2}\sigma^2\frac{1}{n}\sum_{i\neq j}\phi_{ij}\right)\\ &= 4\sigma^4 n^{-2}|\mathbf{H}|\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}\mathbb{C}\mathrm{ov}(\phi_{ij},\phi_{kl}). \quad (2.35)\end{aligned}$$

Now, consider the value of $\mathbb{C}\mathrm{ov}(\phi_{ij},\phi_{kl})$ according to the following three exclusive cases: when $i=k$ and $j=l$, when $i=k$ and $j\neq l$, and, finally, when i,j,k,l are all distinct.

First, when $i=k$ and $j=l$, the total number of such terms is $n(n-1)/2$. In this case, it follows that

$$\begin{aligned}\mathbb{C}\mathrm{ov}(\phi_{ij},\phi_{ij}) &\leq \mathbb{E}\Big[\iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{t}\right)K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\\ &\quad \cdot K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)\rho_n^2(\mathbf{X}_i-\mathbf{X}_j)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\Big]\\ &= \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\Big\{\iint K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{t}\right)K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{x}\right)\\ &\quad \cdot K_{\mathbf{H}}\left(\mathbf{v}-\mathbf{t}\right)\rho_n^2(\mathbf{u}-\mathbf{v})f(\mathbf{u})f(\mathbf{v})d\mathbf{u}d\mathbf{v}\Big\}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\ &= \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\Big\{|\mathbf{H}|^{-2}\iint K[-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]\\ &\quad \cdot K[-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)f(\mathbf{t}+\mathbf{H}\mathbf{p})f(\mathbf{t}+\mathbf{H}\mathbf{q})\\ &\quad \cdot\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}\Big\}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\ &= \iint \frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\Big\{|\mathbf{H}|^{-2}f^2(\mathbf{t})\iint K[-\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\\ &\quad \cdot K[-\mathbf{q}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}+o(1)\Big\}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\ &= |\mathbf{H}|^{-1}\iiiint \frac{1}{f^2(\mathbf{x})}K\left(-\mathbf{p}+\mathbf{u}\right)K\left(-\mathbf{q}+\mathbf{u}\right)K\left(\mathbf{p}\right)K\left(\mathbf{q}\right)\\ &\quad w(\mathbf{x})w(\mathbf{x}-\mathbf{H}\mathbf{u})\cdot\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}d\mathbf{x}d\mathbf{u}\cdot[1+o(1)]\\ &\leq \frac{K_M^4 w_M^2}{f_M^2 n|\mathbf{H}|^2}\int_{\|\mathbf{q}\|\leq 1}\{n|\mathbf{H}|\int_{\|\mathbf{p}\|\leq 1}\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}\}d\mathbf{q}\cdot[1+o(1)].\end{aligned}$$

Since

$$n|\mathbf{H}|\int_{\|\mathbf{p}\|\le 1}\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}\le n\int|\rho_n(\mathbf{t})|d\mathbf{t}\le C,$$

then

$$\mathbb{C}\mathrm{ov}(\phi_{ij},\phi_{ij}) \le \frac{K_M^4 w_M^2}{f_M^2}\frac{C}{n|\mathbf{H}|^2}\times \text{Volume of the ball } \{\mathbf{u}:\|\mathbf{u}\|\le 1\} = \frac{C_2}{n|\mathbf{H}|^2}. \quad (2.36)$$

Second, when $i=k$ and $j\neq l$ in the expression (2.35), the total number of such terms can be bounded by n^3, and it follows that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}(\phi_{ij},\phi_{il}) &= \mathbb{E}(\phi_{ij}\phi_{il})-\mathbb{E}(\phi_{ij})\mathbb{E}(\phi_{il}) = \mathbb{E}(\phi_{ij}\phi_{il})-[\mathbb{E}(\phi_{ij})]^2 \le \mathbb{E}(\phi_{ij}\phi_{il})\\
&\le \mathbb{E}\Big[\iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{t})\,K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\\
&\quad\cdot K_{\mathbf{H}}(\mathbf{X}_l-\mathbf{t})\,\rho_n(\mathbf{X}_i-\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_l)w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\Big]\\
&= \iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}\iiint K_{\mathbf{H}}(\mathbf{u}-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{v}-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{u}-\mathbf{t})\\
&\quad\cdot K_{\mathbf{H}}(\mathbf{y}-\mathbf{t})\,\rho_n(\mathbf{u}-\mathbf{y})\rho_n(\mathbf{u}-\mathbf{v})f(\mathbf{u})f(\mathbf{v})f(\mathbf{y})d\mathbf{u}d\mathbf{v}d\mathbf{y}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= \iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}|\mathbf{H}|^{-1}\iiint K(\mathbf{p})K(\mathbf{q})\,K[\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]\\
&\quad\cdot K(\mathbf{r})\,f(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{q})f(\mathbf{t}+\mathbf{H}\mathbf{r})\\
&\quad\cdot\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})]d\mathbf{p}d\mathbf{q}d\mathbf{r}w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= \iint\frac{1}{f^2(\mathbf{x})f^2(\mathbf{t})}|\mathbf{H}|^{-1}f^2(\mathbf{x})f(\mathbf{t})\iiint K(\mathbf{p})K(\mathbf{q})\\
&\quad\cdot K[\mathbf{p}+\mathbf{H}^{-1}(\mathbf{x}-\mathbf{t})]K(\mathbf{r})\,\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]\rho_n[\mathbf{x}-\mathbf{t}+\mathbf{H}(\mathbf{p}-\mathbf{r})]d\mathbf{p}d\mathbf{q}d\mathbf{r}\\
&\quad\cdot [1+o(1)]w(\mathbf{x})w(\mathbf{t})d\mathbf{x}d\mathbf{t}\\
&= \int\cdots\int\frac{1}{f(\mathbf{t})}K(\mathbf{p})\,K(\mathbf{p}+\mathbf{u})K(\mathbf{q})\,K(\mathbf{r})\,w(\mathbf{t}+\mathbf{H}\mathbf{u})w(\mathbf{t})\\
&\quad\cdot\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})]\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q}d\mathbf{r}d\mathbf{u}d\mathbf{t}\cdot[1+o(1)]\\
&= n^{-2}|\mathbf{H}|^{-2}\iiint\frac{1}{f(\mathbf{t})}K(\mathbf{p})\,K(\mathbf{p}+\mathbf{u})\,w^2(\mathbf{t})\\
&\quad\cdot\{n|\mathbf{H}|\int K(\mathbf{r})\,\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})]d\mathbf{r}\}\\
&\quad\cdot\{n|\mathbf{H}|\int K(\mathbf{q})\,\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{q}\}d\mathbf{p}d\mathbf{u}d\mathbf{t}\cdot[1+o(1)].
\end{aligned}$$

Since

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{r})\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r}+\mathbf{u})]d\mathbf{r} = K(\mathbf{p}+\mathbf{u})\rho_c,$$

and

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{q})\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{q} = K(\mathbf{p})\rho_c,$$

and taking into account that the functions K, and w are bounded, and f is bounded away from zero, it is obtained that

$$\mathbb{C}\mathrm{ov}(\phi_{ij}, \phi_{il}) \leq \frac{C_3}{n^2|\mathbf{H}|^2}. \tag{2.37}$$

Finally, when i, j, k, l are all distinct in (2.35), given that ϕ_{ij} and ϕ_{kl} are independent,

$$\mathbb{C}\mathrm{ov}(\phi_{ij}, \phi_{kl}) = 0. \tag{2.38}$$

Then, considering (2.35), (2.36), (2.37) and (2.38), it follows that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\mathbb{E}(I_{32}|\mathbf{X}_1,\ldots,\mathbf{X}_n)] &= \mathbb{V}\mathrm{ar}\left(|\mathbf{H}|^{1/2}\sigma^2\frac{1}{n}\sum_{i\neq j}\phi_{ij}\right) \\
&\leq 4\sigma^4 n^{-2}|\mathbf{H}|\left(\frac{n^2-n}{2}\frac{C_2}{n|\mathbf{H}|^2} + n^3\frac{C_3}{n^2|\mathbf{H}|^2}\right) \\
&= \mathcal{O}_{\mathbb{P}}(n^{-1}|\mathbf{H}|^{-1}).
\end{aligned} \tag{2.39}$$

Now, from (2.23), (2.34) and (2.39), the leading term of the variance of I_{32} is given by:

$$V = 2\sigma^4 K^{[4]}(\mathbf{0})\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c\int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2\int w^2(\mathbf{x})d\mathbf{x}\right]. \tag{2.40}$$

Therefore, using the central limit theorem for degenerate reduced U-statistics under α-mixing conditions, given by Kim et al. (2013), it is obtained that the term I_{32} converges in distribution to a normal distribution with mean the leading term of (2.22) and variance given by (2.40).

On the other hand, in virtue of the Cauchy–Schwarz inequality, the cross terms in T_{n1} resulting from the products of $I_1(\mathbf{x})$, $I_2(\mathbf{x})$ and $I_3(\mathbf{x})$ are all of smaller order.

Therefore, combining the results in (2.13), (2.14) and (2.20), and the asymptotic normality of I_{32} (with bias the leading term of (2.22) and variance (2.40)), one gets

$$V^{-1/2}(T_{n1} - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty, \tag{2.41}$$

where

$$\begin{aligned} b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2}\sigma^2 K^{[2]}(\mathbf{0})\left[\int \frac{w(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + \rho_c \int w(\mathbf{x})d\mathbf{x}\right], \\ b_{1\mathbf{H}} &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x}, \\ V &= 2\sigma^4 K^{[4]}(\mathbf{0})\left[\int \frac{w^2(\mathbf{x})}{f^2(\mathbf{x})}d\mathbf{x} + 2\rho_c \int \frac{w^2(\mathbf{x})}{f(\mathbf{x})}d\mathbf{x} + 4\rho_c^2 \int w^2(\mathbf{x})d\mathbf{x}\right]. \end{aligned}$$

The term T_{n2} in T_n is of smaller order than T_{n1} (specifically, $T_{n2} = \mathcal{O}_{\mathbb{P}}[\text{tr}(\mathbf{H}^2)T_{n1}]$), and by the Cauchy-Schwarz inequality, the cross term T_{n12} is of smaller order as well. Therefore, from (2.12), it follows that

$$T_n = T_{n1} + \mathcal{O}_{\mathbb{P}}[\text{tr}(\mathbf{H}^2)] + \mathcal{O}_{\mathbb{P}}(n^{\frac{-2}{4+d}+\eta}).$$

Taking into account (2.41), it follows that

$$V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,$$

with $b_{0\mathbf{H}}$, $b_{1\mathbf{H}}$ and V given above.

Proof of Theorem 2.2 The proof of this result is provided using the Nadaraya–Watson estimator, that is, when $p = 0$ in (2.1). For simplicity, in this proof, $T_{n,0}$ will be denoted by T_n. The test statistic (2.1) can be decomposed as:

$$\begin{aligned} T_n &= n|\mathbf{H}|^{1/2} \int [\hat{m}_{\mathbf{H}}(\mathbf{x};0) - \hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};0)]^2 w(\mathbf{x})d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \left[\frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})Z_i}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})} - \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}\right]^2 w(\mathbf{x})d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \frac{\left\{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\left[m(\mathbf{x}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)\right]\right\}^2}{\left[\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\right]^2} w(\mathbf{x})d\mathbf{x}. \end{aligned}$$

Now, taking into account that the trend considered are of the form $m = m_{\boldsymbol{\beta}_0} + n^{-1/2}|\mathbf{H}|^{-1/4}g$ (see Section 2.2 for further details), one gets

$$\begin{aligned} T_n &= n|\mathbf{H}|^{1/2} \int \frac{\left\{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\left[m_{\boldsymbol{\beta}_0}(\mathbf{x}_i) + n^{-1/2}|\mathbf{H}|^{-1/4}g(\mathbf{x}_i) + \varepsilon_i - m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)\right]\right\}^2}{\left[\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\right]^2} \\ &\quad \cdot w(\mathbf{x})d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int [I_1(\mathbf{x}) + I_2(\mathbf{x}) + I_3(\mathbf{x})]^2 w(\mathbf{x})d\mathbf{x}, \end{aligned}$$

where

$$\begin{aligned} I_1(\mathbf{x}) &= \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\left[m_{\boldsymbol{\beta}_0}(\mathbf{x}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)\right]}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}, \\ I_2(\mathbf{x}) &= \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\, n^{-1/2}|\mathbf{H}|^{-1/4}g(\mathbf{x}_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}, \\ I_3(\mathbf{x}) &= \frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\,\varepsilon_i}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}. \end{aligned}$$

For the term $I_1(\mathbf{x})$, Riemann approximations of sums by integrals can be employed to compute the approximation of $\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\left[m_{\boldsymbol{\beta}_0}(\mathbf{x}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)\right]$ by an integral. For this, an analogous procedure to that used in the proof of Lemma A.4 can be used. Under assumptions (A2), (A5) and (K1), using this type of approximations, and given that the difference $m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}) - m_{\boldsymbol{\beta}_0}(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}(n^{-1/2})$ uniformly in $\mathbf{x}$, it is obtained that

$$\begin{aligned} & n|\mathbf{H}|^{1/2} \int I_1^2(\mathbf{x})w(\mathbf{x})d\mathbf{x} \\ &= n|\mathbf{H}|^{1/2} \int \left\{\frac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})\left[m_{\boldsymbol{\beta}_0}(\mathbf{x}_i) - m_{\hat{\boldsymbol{\beta}}}(\mathbf{x}_i)\right]}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x})}\right\}^2 w(\mathbf{x})d\mathbf{x} \\ &= \mathcal{O}_{\mathbb{P}}(|\mathbf{H}|^{1/2}). \end{aligned} \tag{2.42}$$

Regarding the terms $I_2(\mathbf{x})$ and $I_3(\mathbf{x})$, following similar arguments to those used in the proof of Theorem 2.1 for these terms, but employing Riemann approximations of sums by integrals (such as the one given in Lemma A.4), it is obtained that

$$
\begin{aligned}
n|\mathbf{H}|^{1/2} \int I_2^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x}, \qquad (2.43)\\
n|\mathbf{H}|^{1/2} \int I_3^2(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} &= I_{31} + I_{32},
\end{aligned}
$$

with

$$
I_{31} = \sigma^2 |\mathbf{H}|^{-1/2} K^{[2]}(\mathbf{0}) \int w(\mathbf{x}) d\mathbf{x} \cdot [1 + o_{\mathbb{P}}(1)], \qquad (2.44)
$$

and I_{32} converging in distribution to a normally distributed random variable with bias

$$
\mathbb{E}(|\mathbf{H}|^{1/2} I_{32}) = \sigma^2 K^{[2]}(\mathbf{0}) \rho_c \int w(\mathbf{x}) d\mathbf{x} \cdot [1 + o(1)]. \qquad (2.45)
$$

and variance

$$
V = \sigma^4 K^{[4]}(\mathbf{0}) \int w^2(\mathbf{x}) d\mathbf{x} (1 + \rho_c + 2\rho_c^2). \qquad (2.46)
$$

In virtue of the Cauchy–Bunyakovsky–Schwarz inequality, the cross terms in T_n resulting from the products of I_1, I_2 and I_3 are all of small order. Therefore, combining the results given in the equations (2.42), (2.43) and (2.44), and the asymptotic normality of I_{32} (with its bias (2.45) and its variance (2.46)), it follows that

$$
V^{-1/2}(T_n - b_{0\mathbf{H}} - b_{1\mathbf{H}}) \to_{\mathcal{L}} N(0,1) \text{ as } n \to \infty,
$$

where

$$
\begin{aligned}
b_{0\mathbf{H}} &= |\mathbf{H}|^{-1/2} \sigma^2 K^{[2]}(\mathbf{0}) \int w(\mathbf{x}) d\mathbf{x} (1 + \rho_c),\\
b_{1\mathbf{H}} &= \int [K_{\mathbf{H}} * g(\mathbf{x})]^2 w(\mathbf{x}) d\mathbf{x},\\
V &= \sigma^4 K^{[4]}(\mathbf{0}) \int w^2(\mathbf{x}) d\mathbf{x} (1 + \rho_c + 2\rho_c^2).
\end{aligned}
$$

Chapter 3

Nonparametric regression estimation for a circular response and an $\mathbb{R}^d$-valued covariate

3.1 Introduction

New challenges on regression modeling appear when trying to describe relations between variables and some of them (response and/or covariates) do not belong to an Euclidean space. This is the case for regression models, where some or all of the involved variables are circular ones. As pointed out in Section 1.3, the special nature of circular data (points on the unit circle; angles in $\mathbb{T} = [0, 2\pi)$) relies on their periodicity, which requires *ad hoc* statistical methods to analyze them. Circular statistics is an evolving discipline, and several statistical techniques for linear data now may claim their circular analogues. Comprehensive reviews on circular statistics (or more general, directional data) are provided by Fisher (1995), Jammalamadaka and SenGupta (2001) or Mardia and Jupp (2000). Some recent advances in directional statistics are collected in Ley and Verdebout (2017). Examples of circular data arise in many scientific fields such as biology, studying animal orientation (see the classical book by Batschelet, 1981), environmental applications (see SenGupta and Ugwuowo, 2006), or oceanography (as in Wang et al., 2015, among others). When the circular variable is supposed to vary with respect to other covariates and the

goal is to model such a relation, regression estimators for circular responses must be designed and analyzed. Parametric regression approaches were originally considered by Fisher and Lee (1992) and Presnell et al. (1998), assuming a parametric (conditional) distribution model for the circular response variable. In this scenario, Euclidean covariates are supposed to influence the response via the parameters of the conditional distribution (e.g. through the location parameter, as the simplest case, or through location and concentration, if a von Mises distribution is chosen). Following the proposal by Presnell et al. (1998), Scapini et al. (2002) analyzed the orientation of two species of sand hoppers, considering parametric multiple regression methods for circular responses. A parametric multivariate circular regression problem was also studied by Kim and SenGupta (2017). Beyond parametric restrictions, flexible approaches are also feasible in this context (circular response and covariates), just imposing some regularity conditions on the regression function, but avoiding the assumption of a specific parametric family neither for the regression function nor for the conditional distribution. Local estimators of the regression function for circular response and a single real-valued covariate were introduced by Di Marzio et al. (2013). The authors proposed nonparametric estimators for the regression function which are defined as the inverse tangent function of the ratio between two sample statistics, obtained as weighted average of the sines and the cosines of the response variable, respectively. Different weights provide alternative estimators. A multivariate angular regression model for both angular and linear predictors was studied by Rivest et al. (2016). Maximum likelihood estimators for the parameters were derived under two von Mises error structures.

The problem of nonparametrically estimating the conditional mean direction of a circular random variable, given an $\mathbb{R}^d$-valued covariate, is considered in this chapter. If the relation between both variables is viewed from a model-based approach, then the proposal aims to estimate the usual target regression function, given by the inverse tangent function of the ratio between the conditional expectations of the sine and cosine of the response variable. The proposal considers two regression models for the sine and cosine components, which are indeed regression models with real-valued response and d-dimensional covariate. Then, nonparametric estimators for the circular regression function are obtained by computing the inverse tangent

function of the ratio of multivariate nonparametric estimators for the two regression functions of the sine and cosine models. The estimators obtained with this proposal generalize to both higher dimensions and higher polynomial degrees the proposals in Di Marzio et al. (2013). The approach of considering two flexible regression models for the sine and cosine components has been also explored in Jammalamadaka and Sarma (1993), where the objective is the estimation of the regression function in a model with circular response and circular covariate. In this case, the conditional expectations of the sine and the cosine of the response are approximated by trigonometric polynomials of a suitable degree. A similar approach has been also considered in Di Marzio et al. (2014), where the problem of nonparametrically estimating a regression function with spherical response and spherical covariate is addressed as a multi-output regression problem. In this case, each Cartesian coordinate of the spherical regression function is separately estimated.

This chapter is organized as follows. Section 3.2 contains a brief review on nonparametric regression estimation considering a regression model with a circular response and a single-real valued covariate. A multiple linear-circular regression model is presented in Section 3.3. In Section 3.3.1, assuming this model, nonparametric estimators of the circular regression function, based on considering two regression models for the sine and cosine components of the response variable, are presented. Their asymptotic biases and variances are derived. Section 3.3.2 presents an alternative formulation for the nonparametric estimators of the circular regression function, by considering weighted average smoothers of the sines and the cosines of the response variable. Their asymptotic properties are also derived. The finite sample performance of the estimators is assessed through a simulation study, provided in Section 3.3.3. Section 3.3.4 shows a real data application about sand hoppers orientation. Finally, Section 3.4 includes the proofs of the main results.

3.2 A brief background on nonparametric circular regression estimation with a single covariate

Let $\{(X_i, \Theta_i)\}_{i=1}^n$ be a random sample from (X, Θ), where Θ is a circular random variable taking values on $\mathbb{T} = [0, 2\pi)$, and X is a real-valued variable with density f

supported on $\mathcal{D} \subseteq \mathbb{R}$. Assume the following regression model:

$$\Theta_i = [m^c(X_i) + \varepsilon_i](\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n, \tag{3.1}$$

where m^c is a circular regression function, and the ε_i are independent and identically distributed (i.i.d.) random angles, with zero mean direction, finite concentration, and independent of the X_i.

The circular regression function m^c in model (3.1) can be defined as the minimizer of the risk $\mathbb{E}\{1-\cos[\Theta-m^c(X)] \mid X = x\}$. It can be proved that the minimizer of this cosine risk is given by $m^c(x) = \text{atan2}[m_1(x), m_2(x)]$, where $m_1(x) = \mathbb{E}[\sin(\Theta) \mid X = x]$, $m_2(x) = \mathbb{E}[\cos(\Theta) \mid X = x]$ and the function atan2 was defined in Section 1.3. Di Marzio et al. (2013) proposed a kernel-type estimator of this circular regression function. Defining $g_1(x) = m_1(x)f(x)$ and $g_2(x) = m_2(x)f(x)$, the minimizer of the cosine risk can be also written as $m^c(x) = \text{atan2}[g_1(x), g_2(x)]$. Consequently, the estimator proposed by Di Marzio et al. (2013) is:

$$\tilde{m}^c_h(x; p) = \text{atan2}[\hat{g}_{1,h}(x; p), \hat{g}_{2,h}(x; p)], \tag{3.2}$$

where, for $p = 0, 1$, $\hat{g}_{1,h}(x; p)$ and $\hat{g}_{2,h}(x; p)$ are defined as follows:

$$\hat{g}_{1,h}(x; p) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_h(X_i - x; p)\sin(\Theta_i) \quad \text{and} \quad \hat{g}_{2,h}(x; p) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_h(X_i - x; p)\cos(\Theta_i),$$

being $\hat{W}_h(u; p)$ a local weight such that $\hat{g}_{1,h}(x; p)/\hat{g}_{2,h}(x; p)$ is asymptotically unbiased for $g_1(x)/g_2(x)$. Di Marzio et al. (2013) considered the weights

$$\hat{W}_h(u; 0) = K_h(u)$$

and

$$\hat{W}_h(u; 1) = \frac{1}{n}K_h(u)\left[\sum_{j=1}^{n} K_h(X_j - x)(X_j - x)^2 - u\sum_{j=1}^{n} K_h(X_j - x)(X_j - x)\right].$$

Note that $\hat{W}_h(u; 0)$ and $\hat{W}_h(u; 1)$ are the addends of the numerator of the Nadaraya–

Watson and the local linear estimators, respectively.

As pointed out in Section 3.1, asymptotic properties of the estimator given in (3.2) were derived for $p = 0$ and $p = 1$ in Theorems 3 and 4 of Di Marzio et al. (2013), respectively.

3.3 Nonparametric circular regression estimation with several covariates

This section is devoted to present nonparametric estimators for a regression model with a circular response and an $\mathbb{R}^d$-valued predictor. Let $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample from $(\mathbf{X}, \Theta)$, where Θ is a circular random variable taking values on $\mathbb{T} = [0, 2\pi)$, and $\mathbf{X}$ is a random variable with density f supported on $\mathcal{D} \subseteq \mathbb{R}^d$. Assume that Θ and $\mathbf{X}$ are related through the following regression model:

$$\Theta_i = [m^c(\mathbf{X}_i) + \varepsilon_i](\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n, \tag{3.3}$$

where m^c is a circular regression function, and $\varepsilon_i, i = 1, \ldots, n$, is an independent sample of a circular variable ε, satisfying $\mathbb{E}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] = 0$ and having finite concentration. Additionally, the following notation is used: $\ell(\mathbf{x}) = \mathbb{E}[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$, $\sigma_1^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$, $\sigma_2^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}[\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$ and $\sigma_{12}(\mathbf{x}) = \mathbb{E}[\sin(\varepsilon)\cos(\varepsilon) \mid \mathbf{X} = \mathbf{x}]$.

The circular regression function m^c in model (3.3) is the conditional mean direction of Θ given $\mathbf{X}$. This function can be defined (at a point $\mathbf{x}$) as the minimizer of the risk $\mathbb{E}\{1 - \cos[\Theta - m^c(\mathbf{X})] \mid \mathbf{X} = \mathbf{x}\}$, which is comparable to the L_2 risk in the circular setting. Specifically, the minimizer of this cosine risk is given by $m^c(\mathbf{x}) = \mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]$, where $m_1(\mathbf{x}) = \mathbb{E}[\sin(\Theta) \mid \mathbf{X} = \mathbf{x}]$ and $m_2(\mathbf{x}) = \mathbb{E}[\cos(\Theta) \mid \mathbf{X} = \mathbf{x}]$. Notice that in this framework two regression models for the sine and cosine of Θ on $\mathbf{X}$, are being implicitly considered. In particular, the following regression models for the sine component:

$$\sin(\Theta_i) = m_1(\mathbf{X}_i) + \xi_i \quad i = 1, \ldots, n, \tag{3.4}$$

and for the cosine component:

$$\cos(\Theta_i) = m_2(\mathbf{X}_i) + \zeta_i \quad i = 1, \ldots, n, \tag{3.5}$$

where the ξ_i and the ζ_i, for $i = 1, \ldots, n$, are independent samples of the variables ξ and ζ, respectively, absolutely bounded by 1, satisfying $\mathbb{E}(\xi \mid \mathbf{X} = \mathbf{x}) = \mathbb{E}(\zeta \mid \mathbf{X} = \mathbf{x}) = 0$. Additionally, the following notation is used: $s_1^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}(\xi \mid \mathbf{X} = \mathbf{x})$, $s_2^2(\mathbf{x}) = \mathbb{V}\mathrm{ar}(\zeta \mid \mathbf{X} = \mathbf{x})$ and $c(\mathbf{x}) = \mathbb{E}(\xi\zeta \mid \mathbf{X} = \mathbf{x})$ at every $\mathbf{x} \in \mathcal{D}$. Defining $\nu_1^2(\mathbf{x}) = \mathbb{E}[\sin^2(\Theta)|\mathbf{X} = \mathbf{x}]$ and $\nu_2^2(\mathbf{x}) = \mathbb{E}[\cos^2(\Theta)|\mathbf{X} = \mathbf{x}]$ and $\tilde{c}(\mathbf{x}) = \mathbb{E}[\cos(\Theta)\sin(\Theta)|\mathbf{X} = \mathbf{x}]$, it is clear that $\nu_j^2(\mathbf{x}) = m_j^2(\mathbf{x}) + s_j^2(\mathbf{x}), j = 1, 2$ and $\tilde{c}(\mathbf{x}) = m_1(\mathbf{x})m_2(\mathbf{x}) + c(\mathbf{x})$.

Some equations relating certain functions referred to model (3.3), and to models (3.4) and (3.5), can be derived. Using the sine and cosine addition formulas in model (3.3), it follows that, for $i = 1, \ldots, n$,

$$\sin(\Theta_i) = \sin[m^c(\mathbf{X}_i)]\cos(\varepsilon_i) + \cos[m^c(\mathbf{X}_i)]\sin(\varepsilon_i) \tag{3.6}$$

and

$$\cos(\Theta_i) = \cos[m^c(\mathbf{X}_i)]\cos(\varepsilon_i) - \sin[m^c(\mathbf{X}_i)]\sin(\varepsilon_i). \tag{3.7}$$

Hence, defining $f_1(\mathbf{x}) = \sin[m^c(\mathbf{x})]$ and $f_2(\mathbf{x}) = \cos[m^c(\mathbf{x})]$ and applying conditional expectations in (3.6) and (3.7), it holds that

$$m_1(\mathbf{x}) = f_1(\mathbf{x})\ell(\mathbf{x}) \quad \text{and} \quad m_2(\mathbf{x}) = f_2(\mathbf{x})\ell(\mathbf{x}). \tag{3.8}$$

Note that $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ correspond to the normalized versions of $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, respectively. Indeed, taking into account that $f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}) = 1$, it can be easily deduced that $\ell(\mathbf{x}) = [m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})]^{1/2}$. Hence, $\ell(\mathbf{x})$ amounts to the mean resultant length of Θ given $\mathbf{X} = \mathbf{x}$, which also corresponds to the mean resultant length of ε given $\mathbf{X} = \mathbf{x}$, since $\mathbb{E}[\sin(\varepsilon) \mid \mathbf{X} = \mathbf{x}] = 0$ is assumed.

Moreover, equating expressions (3.4) and (3.6), and (3.5) and (3.7), and using (3.8), the errors in models (3.4) and (3.5) can be written as:

$$\xi_i = f_1(\mathbf{X}_i)[\cos(\varepsilon_i) - \ell(\mathbf{X}_i)] + f_2(\mathbf{X}_i)\sin(\varepsilon_i) \quad i = 1, \ldots, n \tag{3.9}$$

and

$$\zeta_i = f_2(\mathbf{X}_i)[\cos(\varepsilon_i) - \ell(\mathbf{X}_i)] - f_1(\mathbf{X}_i)\sin(\varepsilon_i) \quad i = 1, \ldots, n. \tag{3.10}$$

As a consequence, the following explicit expressions for the conditional variances of the error terms involved in models (3.4) and (3.5), can be obtained:

$$s_1^2(\mathbf{x}) = f_1^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_1(\mathbf{x})f_2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^2(\mathbf{x})\sigma_1^2(\mathbf{x}), \tag{3.11}$$

$$s_2^2(\mathbf{x}) = f_2^2(\mathbf{x})\sigma_2^2(\mathbf{x}) - 2f_2(\mathbf{x})f_1(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_1^2(\mathbf{x})\sigma_1^2(\mathbf{x}), \tag{3.12}$$

as well as for the covariance between the error terms in (3.4) and (3.5):

$$c(\mathbf{x}) = f_1(\mathbf{x})f_2(\mathbf{x})\sigma_2^2(\mathbf{x}) - f_1^2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^2(\mathbf{x})\sigma_{12}(\mathbf{x}) - f_1(\mathbf{x})f_2(\mathbf{x})\sigma_1^2(\mathbf{x}). \tag{3.13}$$

As pointed out in Section 3.1, two approaches to derive regression estimators of m^c are designed and the corresponding regression estimators are analyzed. Both approximations account the regression models (3.4) and (3.5) for the sine and cosine components of the response variable. However, the first proposal consists in considering local polynomial type estimators for m_1 and m_2, while the second (equivalent) approach focuses on functions of m_1 and m_2 and considers weighted average smoothers of these functions. This alternative formulation extends in some way the results of Di Marzio et al. (2013) for an arbitrary dimension d.

3.3.1 Local polynomial type estimators

A direct nonparametric regression estimator for model (3.3) is presented and studied in this section. Given that the minimizer of the cosine risk is given by $m^c(\mathbf{x}) = \text{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]$, replacing m_1 and m_2 by appropriate estimators, an estimator for m^c can be directly obtained. In particular, a whole class of kernel-type estimators for m^c at $\mathbf{x} \in \mathcal{D}$ can be defined by considering local polynomial estimators for $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$. Specifically, estimators of the form:

$$\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p) = \text{atan2}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)] \tag{3.14}$$

are considered, where for any integer $p \geq 0$, $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; p)$ denote the pth order local polynomial estimators (with bandwidth matrix $\mathbf{H}$) of $m_1(\mathbf{x})$ and $m_2(\mathbf{x})$, respectively. The special cases $p = 0$ and $p = 1$ yield a Nadaraya–Watson-type estimator and a local linear-type estimator of $m^c(\mathbf{x})$, respectively.

Asymptotic (conditional) bias and variance of the estimator given in (3.14) are derived below. We will mainly focus on the cases in which $p = 0$ and $p = 1$. The asymptotic properties of the corresponding Nadaraya–Watson and local linear estimators of $m_j(\mathbf{x})$, $j = 1, 2$, are firstly recalled just considering that models (3.4) and (3.5) hold. These results are then used to obtain the asymptotic properties of the estimator presented in (3.14) with polynomial degrees $p = 0$ and $p = 1$. Finally, asymptotic properties of local polynomial estimators with a higher order p and $\mathcal{D} \subseteq \mathbb{R}$ are also studied.

Nadaraya–Watson-type estimator

Considering models (3.4) and (3.5), and using (1.6), Nadaraya–Watson estimators for the regression functions m_j, $j = 1, 2$, at a given point $\mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^d$, are respectively defined as:

$$\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0) = \begin{cases} \dfrac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \sin(\Theta_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 1, \\ \dfrac{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \cos(\Theta_i)}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 2. \end{cases} \tag{3.15}$$

Next, the asymptotic conditional bias and variance expressions for $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 0)$ are derived. First, using the asymptotic properties on the multivariate Nadaraya–Watson estimator given in (1.7) and (1.8), the asymptotic conditional bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 0)$, $j = 1, 2$, must be calculated. These preliminary results, along with the covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 0)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; 0)$, are collected in Lemma 3.1. The following assumption is required.

(C1) All second-order derivatives of the regression functions m_j and s_j^2, for $j = 1, 2$, are continuous at $\mathbf{x} \in \mathcal{D}$, and $s_j^2(\mathbf{x}) > 0$.

Lemma 3.1 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *Under assumptions* (A1), (C1), (H1) and (K1), *if* $\mathbf{x}$ *is an interior point of the support of* f, *then, for* $j = 1, 2$,

$$\begin{aligned}
\mathbb{E}[\hat{m}_{j,\mathbf{H}}(\mathbf{x};0) - m_j(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})] \\
&\quad + \frac{\mu_2(K)}{f(\mathbf{x})} \boldsymbol{\nabla}^{\mathrm{T}} m_j(\mathbf{x}) \mathbf{H}^2 \boldsymbol{\nabla} f(\mathbf{x}) \\
&\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \qquad (3.16) \\
\mathbb{V}\mathrm{ar}[\hat{m}_{j,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K) s_j^2(\mathbf{x})}{n|\mathbf{H}| f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \\
\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0), \hat{m}_{2,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K) c(\mathbf{x})}{n|\mathbf{H}| f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

Proof See Section 3.4.

Now, using the previous lemma, the following theorem provides the asymptotic conditional bias and the asymptotic conditional variance of the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$.

Theorem 3.1 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *Under assumptions* (A1), (C1), (H1) and (K1), *the asymptotic conditional bias of estimator* $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$, *at a fixed interior point* $\mathbf{x}$ *in the support of* f, *is given by:*

$$\begin{aligned}
\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) - m^c(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] \\
&\quad + \frac{\mu_2(K)}{\ell(\mathbf{x}) f(\mathbf{x})} \boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x}) \mathbf{H}^2 \boldsymbol{\nabla}(\ell f)(\mathbf{x}) \\
&\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \qquad (3.17)
\end{aligned}$$

and the asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x}) f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \qquad (3.18)$$

Proof See Section 3.4.

Note that both the asymptotic conditional bias and the asymptotic conditional

variance of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$ share the form of the corresponding quantities for the Nadaraya–Watson estimator of a regression function with real-valued response. In the asymptotic bias expression, both the gradient and the Hessian matrix of m^c refer to a circular regression function. In addition, the asymptotic conditional variance depends on the ratio $\sigma_1^2(\mathbf{x})/\ell^2(\mathbf{x})$, accounting for the variability of the errors in model (3.3).

Similarly to the Euclidean case (see Section 1.1.1 for further details), from Theorem 3.1, it is possible to define the AMSE of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$, as the sum of the square of the main term of bias and the main term of the variance,

$$\begin{aligned}
&\mathrm{AMSE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)] \\
= &\left\{\frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x})\right\}^2 + \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} \\
= &\frac{1}{4}\mu_2^2(K)\mathrm{tr}^2\left(\mathbf{H}^2\left\{\frac{1}{\ell(\mathbf{x})f(\mathbf{x})}[\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x}) + \boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x})]\right.\right. \\
&\left.\left.+\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})\right\}\right) + \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}. \qquad (3.19)
\end{aligned}$$

The minimizer of equation (3.19), with respect to $\mathbf{H}$, provides an asymptotically optimal local bandwidth matrix for $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$, which is given by:

$$\mathbf{H}^c_{\mathrm{opt}}(\mathbf{x};0) = \left[\frac{R(K)\sigma_1^2(\mathbf{x})}{nd\mu_2^2(K)f(\mathbf{x})}|\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x})|^{1/2}\right]^{1/(d+4)}\cdot\left[\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x})\right]^{-1/2}, \qquad (3.20)$$

where

$$\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x}) = \begin{cases} \boldsymbol{\mathcal{G}}^c(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{G}}^c(\mathbf{x}) \text{ is positive definite,} \\ -\boldsymbol{\mathcal{G}}^c(\mathbf{x}) & \text{if} \quad \boldsymbol{\mathcal{G}}^c(\mathbf{x}) \text{ is negative definite,} \end{cases} \qquad (3.21)$$

with

$$\boldsymbol{\mathcal{G}}^c(\mathbf{x}) = \frac{1}{\ell(\mathbf{x})f(\mathbf{x})}[\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x}) + \boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x})] + \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}).$$

The previous minimization problem can be solved using Proposition A.3 included

in Appendix A. Notice that in the particular case of $\boldsymbol{H} = h\mathbf{I}_d$, the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 0)$ given in (3.14), with $\mathbf{x}$ being an interior point of the support, achieves an optimal convergence rate of $n^{-4/(d+4)}$, which is the same as the one for the multivariate Nadaraya–Watson estimator with real-valued response (Härdle and Müller, 2012).

Despite deriving the previous explicit expression for the local optimal bandwidth (3.20), its use in practice is limited given that it depends on unknown functions, such as the design density f and the variance of the sine of the errors σ_1^2. In addition, as pointed out in Section 1.1.1, it is more usual in practice to consider a global bandwidth for the whole curve. An asymptotic global optimal bandwidth matrix $\mathbf{H}$ could be obtained by minimizing a global error measurement (such as the integrated version of the AMSE). Again, this will depend on unknowns and, moreover, this optimization problem is not trivial, not being possible to obtain a closed form solution. Alternatively, a suitable adapted cross-validation criterion can be used to select the bandwidth matrix. This is indeed the bandwidth selection method employed in the numerical analysis and the real data application. More details will be provided in Section 3.3.3.

Local linear-type estimator

This section is devoted to present the local linear case, corresponding to $p = 1$. Specifically, for models (3.4) and (3.5), and using (1.9), the local linear estimators of the regression functions m_j, $j = 1, 2$, at $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$, are defined by:

$$\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1) = \begin{cases} \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{S} & \text{if } j = 1, \\ \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\mathbf{C} & \text{if } j = 2, \end{cases} \tag{3.22}$$

where $\mathbf{S} = [\sin(\Theta_1), \ldots, \sin(\Theta_n)]^{\mathrm{T}}$ and $\mathbf{C} = [\cos(\Theta_1), \ldots, \cos(\Theta_n)]^{\mathrm{T}}$.

The asymptotic conditional bias, variance and covariance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1)$, $j = 1, 2$, are calculated. Note that the asymptotic conditional bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; 1)$, $j = 1, 2$, can be directly obtained using the asymptotic results for the multivariate local linear estimator given in (1.10) and (1.11). The expressions, along with the covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 1)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x}; 1)$ are provided in the following result.

Lemma 3.2 *Let $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample from a density supported on $\mathcal{D} \times \mathbb{T}$. Under assumptions* (A1), (C1), (H1) and (K1)*, if $\mathbf{x}$ is an interior point of the support of f, then, for $j = 1, 2$,*

$$\begin{aligned}
\mathbb{E}[\hat{m}_{j,\mathbf{H}}(\mathbf{x};1) - m_j(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})] \\
&\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \qquad (3.23) \\
\mathbb{V}\mathrm{ar}[\hat{m}_{j,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K)s_j^2(\mathbf{x})}{n|\mathbf{H}|f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \\
\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1), \hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K)c(\mathbf{x})}{n|\mathbf{H}|f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

Proof See Section 3.4.

The following theorem provides the asymptotic conditional bias and the asymptotic conditional variance of the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ given in (3.14).

Theorem 3.2 *Let $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ be a random sample from a density supported on $\mathcal{D} \times \mathbb{T}$. Under assumptions* (A1), (C1), (H1) and (K1)*, the asymptotic conditional bias of estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$, with $\mathbf{x}$ being a fixed interior point in the support of f, is given by:*

$$\begin{aligned}
\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1) - m^c(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2 \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] \\
&\quad + \frac{\mu_2(K)}{\ell(\mathbf{x})} \boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x}) \mathbf{H}^2 \boldsymbol{\nabla} \ell(\mathbf{x}) \\
&\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)], \qquad (3.24)
\end{aligned}$$

while its asymptotic conditional variance is:

$$\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \qquad (3.25)$$

Proof See Section 3.4.

Estimators $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$ and $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ have the same leading terms in their asymptotic conditional variances, while their asymptotic conditional biases, also being of the same order, have different leading terms. In particular, the main term of the

asymptotic conditional bias of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ does not depend on the density function f. Moreover, as a consequence of its definition, the local linear-type estimator, differently from the Nadaraya–Watson-type one, automatically adapts to boundary regions, in the sense that for compactly supported f, the asymptotic conditional bias has the same order both for the interior and for the boundary of the support of f (see Ruppert and Wand, 1994).

As a consequence of Theorem 3.2, an asymptotically optimal local bandwidth can be also obtained for $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$, which coincides with (3.20), but taking $\boldsymbol{\mathcal{G}}^c(\mathbf{x}) = \ell^{-1}(\mathbf{x})[\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x}) + \boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}\ell(\mathbf{x})] + \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})$.

Higher order polynomials for univariate covariate

Asymptotic theory on local polynomial estimators (Fan and Gijbels, 1996) can be used to generalize the above results to local polynomial estimators of arbitrary order p. Here, following the lines in Ruppert and Wand (1994), we will only focus on the case $d = 1$ to analyze asymptotically the nonparametric regression estimator given in (3.14) for $p > 1$. In particular, from (1.13), the pth degree local polynomial estimators for m_j, $j = 1, 2$, at $x \in \mathcal{D} \subseteq \mathbb{R}$, are:

$$\hat{m}_{j,h}(x;p) = \begin{cases} \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_x\boldsymbol{\mathcal{X}}_{x,p})^{-1}\boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_x\mathbf{S} & \text{if } j = 1, \\ \\ \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_x\boldsymbol{\mathcal{X}}_{x,p})^{-1}\boldsymbol{\mathcal{X}}_{x,p}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_x\mathbf{C} & \text{if } j = 2. \end{cases} \tag{3.26}$$

In this univariate framework, the pth degree local polynomial estimator of m^c at x, denoted by $\hat{m}^c_h(x;p)$, has the same expression as the one given in (3.14), but using estimators $\hat{m}_{j,h}(x;p)$, $j = 1, 2$, defined in (3.26), as the arguments of the atan2 function. Considering (1.14), (1.15) and (1.16), it is clear that the conditional asymptotic bias of $\hat{m}^c_h(x;p)$ will depend on whether the polynomial degree is even or odd. Since computations are tedious for high-order polynomials, asymptotic properties of estimator $\hat{m}^c_h(x;p)$ at $x \in \mathcal{D} \subset \mathbb{R}^d$ will be derived only when the polynomial degree p is equal to two and three. Notice that in the case of the regression function, Fan and Gijbels (1996) recommend to use polynomial orders $p = 1$ or $p = 3$ for estimating this curve. Results could be extended for higher-order polynomial degrees.

Theorem 3.3 *Let $\{(X_i, \Theta_i)\}_{i=1}^n$ be a random sample from (X, Θ) supported on $\mathcal{D} \times \mathbb{T}$, with $\mathcal{D} \subseteq \mathbb{R}$, and let x be an interior point of the support of the design density f. Under assumptions* (A1), (C1), (H1) and (K1) *(adapted for $d = 1$) and assuming that m_j, $j = 1, 2$, admits continuous derivatives up to order four in a neighborhood of x, then,*

$$\begin{aligned}\mathbb{E}[\hat{m}_h^c(x;2) - m^c(x) \mid X_1, \ldots, X_n] &= \frac{h^4 \mu_4(K_{(2)}) f^{(1)}(x)}{3! f(x)} [m^{c(3)}(x) + a(x)] \\ &\quad + \frac{h^4 \mu_4(K_{(2)})}{4!} [m^{c\,(4)}(x) + b(x)] + o_{\mathbb{P}}\left(h^4\right)\end{aligned}$$

and

$$\mathbb{V}\mathrm{ar}[\hat{m}_h^c(x;2) \mid X_1, \ldots, X_n] = \frac{R\left(K_{(2)}\right)}{nh\ell^2(x) f(x)} \sigma_1^2(x) + o_{\mathbb{P}}\left(\frac{1}{nh}\right),$$

where

$$\begin{aligned}a(x) &= \frac{2\ell^{(2)}(x) m^{c(1)}(x) + 4\ell^{(1)}(x) m^{c(2)}(x)}{\ell(x)} \\ &\quad + \frac{m_2^{(2)}(x) m_1^{(1)}(x) - m_1^{(2)}(x) m_2^{(1)}(x) + 2{\ell^{(1)}}^2(x) m^{c(1)}(x)}{\ell^2(x)}\end{aligned}$$

and

$$\begin{aligned}b(x) &= \frac{2\ell^{(3)}(x) m^{c(1)}(x) + 6\ell^{(1)}(x) m^{c(3)}(x) + 6\ell^{(2)}(x) m^{c(2)}(x)}{\ell(x)} \\ &\quad + \frac{2m_2^{(3)}(x) m_1^{(1)}(x) - 2m_1^{(3)}(x) m_2^{(1)}(x)}{\ell^2(x)} \\ &\quad + \frac{6{\ell^{(1)}}^2(x) m^{c(2)}(x) + 6\ell^{(1)}(x)\ell^{(2)}(x) m^{c(1)}(x)}{\ell^2(x)}.\end{aligned}$$

Proof See Section 3.4.

Theorem 3.4 *Let $\{(X_i, \Theta_i)\}_{i=1}^n$ be a random sample from (X, Θ) supported on $\mathcal{D} \times \mathbb{T}$, with $\mathcal{D} \subseteq \mathbb{R}$, and let x be an interior point of the support of the design density f. Under assumptions* (A1), (C1), (H1) and (K1) *(adapted for $d = 1$) and assuming that m_j, $j = 1, 2$, admits continuous derivatives up to order five in a neighborhood*

of x, then,

$$\mathbb{E}[\hat{m}_h^c(x;3) - m^c(x) \mid X_1, \ldots, X_n] = \frac{h^4 \mu_4(K_{(3)})}{4!}[m^{c(4)}(x) + b(x)] + o_{\mathbb{P}}\left(h^4\right)$$

and

$$\mathbb{V}\mathrm{ar}[\hat{m}_h^c(x;3) \mid X_1, \ldots, X_n] = \frac{R\left(K_{(3)}\right)}{nh\ell^2(x)f(x)}\sigma_1^2(x) + o_{\mathbb{P}}\left(\frac{1}{nh}\right).$$

Proof See Section 3.4.

Similar arguments to those used to prove Theorems 3.1, 3.2, 3.3 and 3.4, can be applied to derive that the conditional bias of the pth order polynomial type estimator, given in (3.14), will be of order $\mathcal{O}_{\mathbb{P}}\{[\mathrm{tr}(\mathbf{H}^2)]^{(p+1)/2}\}$. Moreover, if p is even, f has a continuous derivative in a neighborhood of $\mathbf{x}$ and $\mathbf{x}$ is an interior point of the support of the design density f, then the bias will be of order $\mathcal{O}_{\mathbb{P}}\{[\mathrm{tr}(\mathbf{H}^2)^{p/2+1}]\}$.

3.3.2 Local weighted average type estimators

In this section, another equivalent approach is considered to estimate the circular regression function m^c in model (3.3). This procedure is an extension of that studied in Di Marzio et al. (2013) for several covariates. As pointed out in Section 3.2 for a single covariate, in this case, defining $g_1(\mathbf{x}) = m_1(\mathbf{x})f(\mathbf{x})$ and $g_2(\mathbf{x}) = m_2(\mathbf{x})f(\mathbf{x})$, the minimizer of the cosine risk $\mathbb{E}\{1-\cos[\Theta - m^c(\mathbf{X})] \mid \mathbf{X} = \mathbf{x}\}$ can be also written as $m^c(\mathbf{x}) = \mathrm{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})]$. Therefore, the proposed estimator with this formulation is defined as:

$$\tilde{m}_{\mathbf{H}}^c(\mathbf{x};p) = \mathrm{atan2}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};p), \hat{g}_{2,\mathbf{H}}(\mathbf{x};p)], \tag{3.27}$$

where, for $p = 0, 1$, the local weighted averages $\hat{g}_{1,\mathbf{H}}(\mathbf{x};p)$ and $\hat{g}_{2,\mathbf{H}}(\mathbf{x};p)$ are:

$$\hat{g}_{1,\mathbf{H}}(\mathbf{x};p) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x};p)\sin(\Theta_i) \tag{3.28}$$

and

$$\hat{g}_{2,\mathbf{H}}(\mathbf{x};p) = \frac{1}{n}\sum_{i=1}^{n} \hat{W}_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x};p)\cos(\Theta_i), \tag{3.29}$$

being $\hat{W}_{\mathbf{H}}(\mathbf{u};p)$, a local weight such as $\hat{g}_{1,\mathbf{H}}(\mathbf{x};p)/\hat{g}_{2,\mathbf{H}}(\mathbf{x};p)$ is asymptotically unbiased for $g_1(\mathbf{x})/g_2(\mathbf{x})$. For several covariates, the weights $\hat{W}_{\mathbf{H}}(\mathbf{u};p), p=0,1$ are:

$$\hat{W}_{\mathbf{H}}(\mathbf{u};0) = K_{\mathbf{H}}(\mathbf{u}) \tag{3.30}$$

and

$$\begin{aligned}\hat{W}_{\mathbf{H}}(\mathbf{u};1) &= \frac{1}{n}K_{\mathbf{H}}(\mathbf{u})\Bigg[\sum_{j=1}^{n}K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_j-\mathbf{x})^{\mathrm{T}}(\mathbf{X}_j-\mathbf{x}) \\ &\quad -\mathbf{u}^{\mathrm{T}}\sum_{j=1}^{n}K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_j-\mathbf{x})\Bigg].\end{aligned} \tag{3.31}$$

Taking into account the definition of the atan2 function, it is clear that the estimator given in (3.27) is equivalent to the one proposed in (3.14). Asymptotic properties of the estimator (3.27) will be obtained considering the Nadaraya–Watson and local linear weights, given in (3.30) and (3.31), respectively.

Nadaraya–Watson weights

Asymptotic bias and variance of the estimator (3.27), considering the weights $\hat{W}_{\mathbf{H}}(\mathbf{u};0)$, given in (3.30), are discussed in this section. In order to derive these expressions, first, the asymptotic expectation, variance and covariance of the weighted averages $\hat{g}_{j,\mathbf{H}}(\mathbf{x};0), j=1,2$ defined in (3.28) and (3.29), are obtained.

Lemma 3.3 *Let $\{(\mathbf{X}_i,\Theta_i)\}_{i=1}^n$ be a random sample from a density supported on $\mathcal{D}\times\mathbb{T}$. If $\mathbf{x}$ is an interior point of the support of f, then under assumptions* (A1), (C1), (H1) and (K1)*, for $j=1,2$, it follows*

$$\begin{aligned}\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] &= g_j(\mathbf{x}) + \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x})] + o[\mathrm{tr}(\mathbf{H}^2)], \\ \mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] &= \frac{1}{n|\mathbf{H}|}R(K)\nu_j^2(\mathbf{x})f(\mathbf{x}) + o\left(\frac{1}{n|\mathbf{H}|}\right), \\ \mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0),\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)] &= \frac{1}{n|\mathbf{H}|}R(K)\tilde{c}(\mathbf{x})f(\mathbf{x}) + o\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

Proof See Section 3.4.

Theorem 3.5 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *If* $\mathbf{x}$ *is an interior point of the support of* f, *then under assumptions* (A1), (C1), (H1) and (K1), *it follows*

$$\begin{aligned}\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] &= m^c(\mathbf{x}) + \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) \\ &\quad + o[\mathrm{tr}(\mathbf{H}^2)], \\ \mathbb{V}\mathrm{ar}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] &= \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

Proof See Section 3.4.

Notice that results obtained in Lemma 3.3 and Theorem 3.5 correspond to the multivariate version of those provided in Lemma 3 and Theorem 3 of Di Marzio et al. (2013), respectively, for interior points of the support of f. This correspondence is immediately clear for the asymptotic bias terms. For the asymptotic variance, the equivalence between the expressions can be obtained considering the relations between the variance of the error term in model (3.3) with the variance of the error terms in models (3.4) and (3.5):

$$\frac{m_1^2(\mathbf{x})s_2^2(\mathbf{x}) + m_2^2(\mathbf{x})s_1^2(\mathbf{x}) - 2m_1(\mathbf{x})m_2(\mathbf{x})c(\mathbf{x})}{m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})} = \sigma_1^2(\mathbf{x}). \tag{3.32}$$

Local linear weights

In this section, the asymptotic expressions for the bias and variance of the estimator given in (3.27) are derived using the weights $\hat{W}_{\mathbf{H}}(\mathbf{u};1)$ given in (3.31).

Lemma 3.4 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *If* $\mathbf{x}$ *is an interior point of the support of* f, *then under assumptions* (A1),

(C1), (H1) and (K1), *for* $j = 1, 2$, *it follows*

$$\begin{aligned}
\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x}; 1)] &= \mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})g_j(\mathbf{x}) \\
&\quad +\mu_2^2(K)\frac{1}{2}\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x})] \\
&\quad -\mu_2^2(K)\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_j(\mathbf{x}) + o[\mathrm{tr}(\mathbf{H}^4)], \\
\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x}; 1)] &= \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)\mathrm{tr}^2(\mathbf{H}^2)f^3(\mathbf{x})\nu_j^2(\mathbf{x}) + o\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right], \\
\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 1), \hat{g}_{2,\mathbf{H}}(\mathbf{x}; 1)] &= \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)\mathrm{tr}^2(\mathbf{H}^2)f^3(\mathbf{x})\tilde{c}(\mathbf{x}) + o\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right].
\end{aligned}$$

Proof See Section 3.4.

Theorem 3.6 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *If* $\mathbf{x}$ *is an interior point of the support of* f, *then under assumptions* (A1), (C1), (H1) and (K1), *it follows*

$$\begin{aligned}
\mathbb{E}[\tilde{m}_{\mathbf{H}}^c(\mathbf{x}; 1)] &= m^c(\mathbf{x}) + \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}\ell(\mathbf{x}) \\
&\quad +o[\mathrm{tr}(\mathbf{H}^2)], \\
\mathbb{V}\mathrm{ar}[\tilde{m}_{\mathbf{H}}^c(\mathbf{x}; 1)] &= \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

Proof See Section 3.4.

Similar comments at those given after Theorem 3.5 can also apply for local linear weights. In this case, the asymptotic expressions given in Lemma 3.4 and Theorem 3.6 are an extension to the multivariate case of those provided in Lemma 4 and Theorem 4 of Di Marzio et al. (2013), respectively, for interior points of the support of f.

Despite using different formulations in Section 3.3.1 and 3.3.2, results in Theorems 3.1 and 3.2 for polynomial degrees $p = 0$ and $p = 1$, respectively, coincide with those obtained in Theorems 3.5 and 3.6. However, note that conditional biases and variances are provided in Section 3.3.1, whereas unconditional results are given in Section 3.3.2.

3.3.3 Simulation study

In order to illustrate the performance of the regression estimators proposed in (3.14) (or equivalently, in (3.27)), a simulation study considering different scenarios and model (3.3) is carried out for $d = 2$ (that is, considering a regression model with a circular response and a bidimensional covariate). For each scenario, 500 samples of size n ($n = 64, 100, 225$ and 400) are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, considering the following regression models, for $i = 1, \ldots, n$:

$$M_1 : \Theta_i = [\text{atan2}(6X_{i1}^5 - 2X_{i1}^3 - 1, -2X_{i2}^5 - 3X_{i2} - 1) + \varepsilon_i](\texttt{mod}\, 2\pi), \quad i = 1, \ldots, n$$

$$M_2 : \Theta_i = [\text{acos}(X_{i1}^5 - 1) + \frac{3}{2}\text{asin}(X_{i2}^3 - X_{i2} + 1) + \varepsilon_i](\texttt{mod}\, 2\pi) \quad i = 1, \ldots, n,$$

where $\{(X_{i1}, X_{i2})\}_{i=1}^n$ denotes a sample of the bidimensional covariate $\mathbf{X} = (X_1, X_2)$, and the circular errors ε_i are drawn from a von Mises distribution $vM(0, \kappa)$, with density function given in (1.40), for different values of the concentration ($\kappa = 5, 10$ and 15).

Figure 3.1 shows two realizations of simulated data (model M_1 in top row and model M_2 in bottom row). In both cases, the sample size is $n = 225$. Left plots show the regression functions evaluated in the regularly spaced sample $\{(X_{i1}, X_{i2})\}_{i=1}^n$. Central panels present the random errors generated from a von Mises distribution with zero mean direction and concentration $\kappa = 5$, for model M_1, and $\kappa = 15$, for model M_2. Right panels show the values of the response variables, obtained adding regression functions and circular errors. It can be seen that the errors in the top row, corresponding to $\kappa = 5$, present more variability than the ones generated with $\kappa = 15$.

Numerical and graphical outputs summarize the finite sample performance of Nadaraya–Watson- and local linear-type estimators in the different scenarios. In all cases, the smoothing parameter is chosen by cross-validation, selecting the bandwidth matrix $\mathbf{H}_{\text{CV}^c}$ that minimizes the function:

$$\text{CV}^c(\mathbf{H}) = \sum_{i=1}^{n} \left\{1 - \cos\left[\Theta_i - \hat{m}^c_{\mathbf{H},-i}(\mathbf{X}_i; p)\right]\right\}, \tag{3.33}$$

where $\hat{m}^c_{\mathbf{H},-i}(\mathbf{X}_i; p)$ stands for the circular Nadaraya–Watson-type estimator ($p = 0$) or the circular local linear-type estimator ($p = 1$), computed using all observations

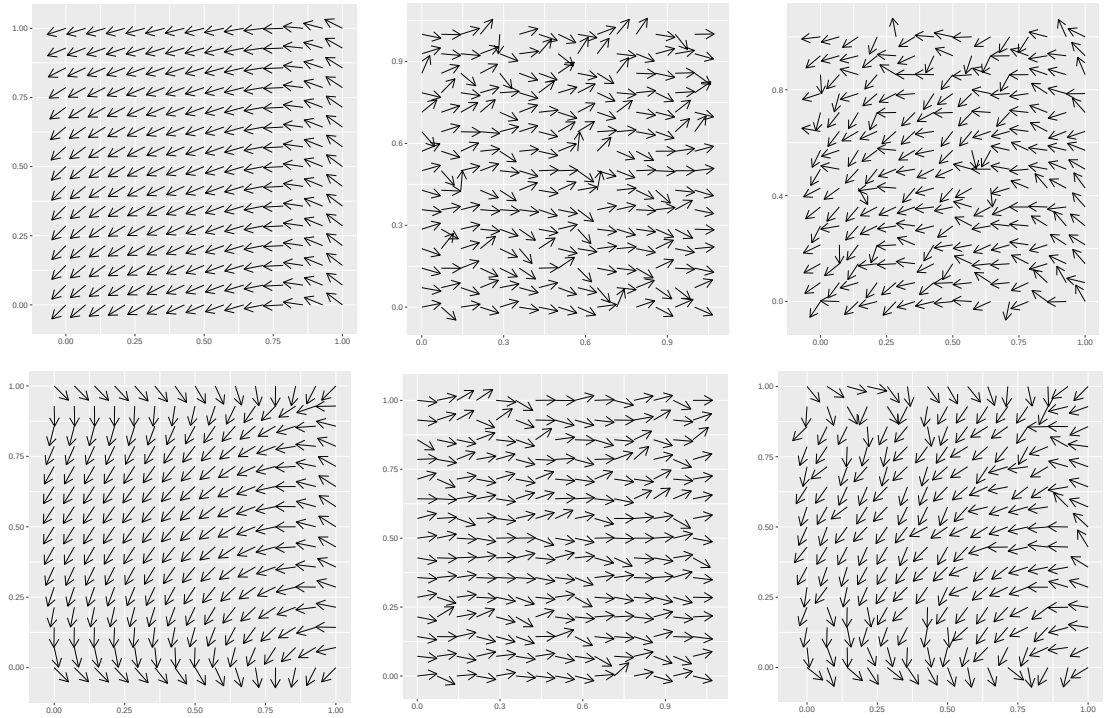

Figure 3.1: Illustration of model generation (model M_1: top row; model M_2: bottom row) on a 15×15 grid. In left panels, regression functions evaluated at the grid points. In center panels, independent errors from a von Mises distribution with zero mean and concentration $\kappa = 5$, for model M_1, and $\kappa = 15$, for model M_2. In right panels, random response variables obtained by adding the two previous plots.

except $(\mathbf{X}_i, \Theta_i)$ and evaluated at $\mathbf{X}_i$. Taking into account the type of regression functions considered in models M_1 and M_2 and to speed up the computing times, in this simulation study, the bandwidth matrix is restricted to be diagonal with possibly different elements. A multivariate triweight kernel is considered for simulations.

Table 3.1 shows, for model M_1 and in the different scenarios, the average (over the 500 replicates) of the circular average squared error (CASE), defined as (Kim and SenGupta, 2017):

$$\text{CASE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)] = \frac{1}{n} \sum_{i=1}^{n} \left\{1 - \cos\left[m^c(\mathbf{X}_i) - \hat{m}^c_{\mathbf{H}}(\mathbf{X}_i; p)\right]\right\}, \tag{3.34}$$

κ	n	Nadaraya–Watson		Local Linear	
		$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{CASE}}$	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{CASE}}$
5	64	0.0610	0.0152	0.0672	0.0147
	100	0.0280	0.0111	0.0358	0.0100
	225	0.0124	0.0066	0.0158	0.0051
	400	0.0075	0.0047	0.0091	0.0033
10	64	0.0094	0.0092	0.0071	0.0076
	100	0.0102	0.0072	0.0055	0.0057
	225	0.0065	0.0042	0.0028	0.0028
	400	0.0042	0.0029	0.0019	0.0019
15	64	0.0182	0.0072	0.0201	0.0056
	100	0.0091	0.0054	0.0110	0.0041
	225	0.0046	0.0032	0.0050	0.0021
	400	0.0032	0.0023	0.0029	0.0014

Table 3.1: Average error (over 500 replicates) of the CASE given in (3.34), for regression model M_1, using Nadaraya–Watson- and local linear-type estimators. Errors are generated from a von Mises distribution with different concentration parameters ($\kappa = 5, 10, 15$). Bandwidth matrix is selected by cross-validation, $\mathbf{H}_{\mathrm{CV}^c}$. Additionally, results when using the optimal bandwidth $\mathbf{H}_{\mathrm{CASE}}$ are also included.

with $p = 0$ (Nadaraya–Watson) and $p = 1$ (local linear), when $\mathbf{H}$ is selected by cross-validation. For comparative purposes, the diagonal optimal bandwidth matrix $\mathbf{H}_{\mathrm{CASE}}$ minimizing (3.34) (obtained by intensive search) is also computed. Note that this bandwidth matrix can not be used in a practical situation where the true regression is unknown. For this reason, it can not be considered as a criterion to select the bandwidth, but it is used to get a benchmark value for comparison. The corresponding averages of the minimum values of the CASE are also included in Table 3.1. It can be seen that the average errors decrease when the sample size increase, and it is smaller for the local linear-type estimator for most of the scenarios. Additionally, as expected, results are generally better when the error concentration gets larger. Average errors of the CASE for model M_2 are shown in Table 3.2.

Numerical outputs are completed with some additional plots. As an illustration of the correct performance of Nadaraya–Watson- and local linear-type estimators, Figure 3.2 shows the theoretical regression functions for models M_1 and M_2 (left

κ	n	Nadaraya–Watson		Local Linear	
		$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{CASE}}$	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{CASE}}$
5	64	0.0638	0.0303	0.0684	0.0209
	100	0.0330	0.0239	0.0369	0.0154
	225	0.0190	0.0158	0.0170	0.0089
	400	0.0141	0.0120	0.0102	0.0061
10	64	0.0297	0.0139	0.0315	0.0088
	100	0.0181	0.0116	0.0172	0.0068
	225	0.0131	0.0087	0.0085	0.0041
	400	0.0109	0.0075	0.0054	0.0029
15	64	0.0198	0.0139	0.0206	0.0088
	100	0.0138	0.0116	0.0118	0.0068
	225	0.0114	0.0087	0.0061	0.0041
	400	0.0100	0.0075	0.0041	0.0029

Table 3.2: Average error (over 500 replicates) of the CASE given in (3.34), for regression model M_2, using Nadaraya–Watson- and local linear-type estimators. Errors are generated from a von Mises distribution with different concentration parameters ($\kappa = 5, 10, 15$). Bandwidth matrix is selected by cross-validation, $\mathbf{H}_{\mathrm{CV}^c}$. Additionally, results when using the optimal bandwidth $\mathbf{H}_{\mathrm{CASE}}$ are also included.

panels) and the corresponding average, over the 500 replicates, of the estimates using the specific scenarios in Figure 3.1 (Nadaraya–Watson- and local linear-type estimates in the center and right panels, respectively). Notice that, for comparison purposes, the theoretical regression functions are plotted in a 100×100 regular grid of the explanatory variables (the same grid where the estimations were computed). Plots in the top row present the results for the data generated from model M_1 and those in the bottom row for model M_2. Although both estimators have a similar and correct behavior, the local linear-type estimator seems to show a slightly better performance, at least, for these samples.

More reliable comparisons between Nadaraya–Watson- and local linear-type estimators can be performed computing the circular bias (CB), the circular variance (CVAR), and the circular mean squared error (CMSE) for both estimators, in a grid

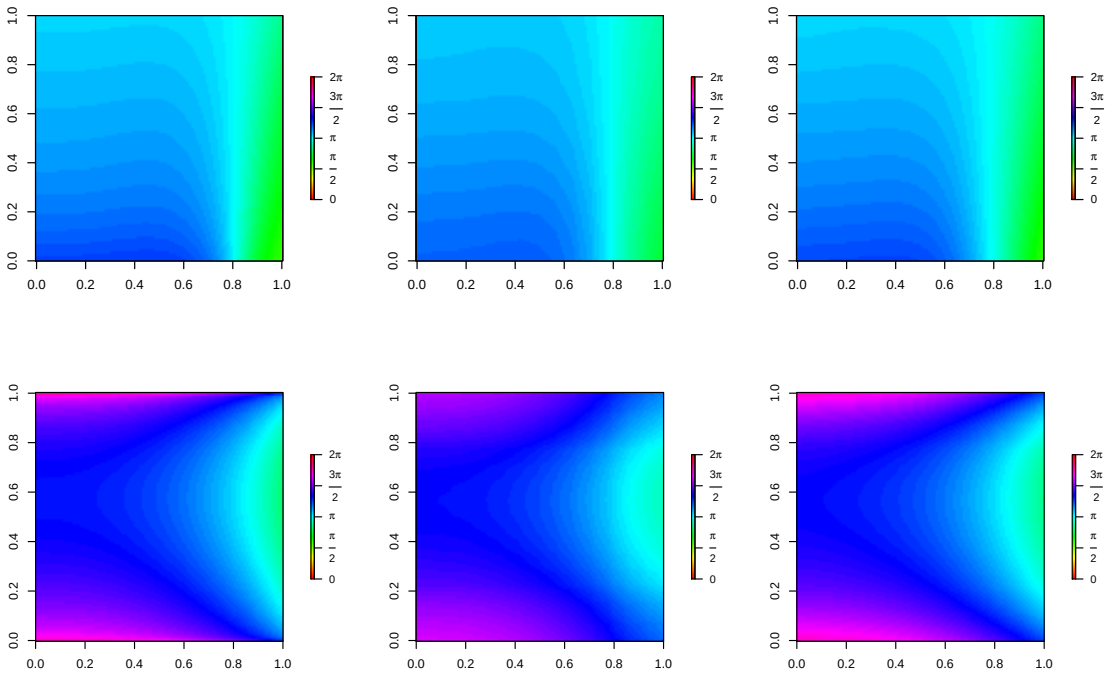

Figure 3.2: Theoretical regression function (left panels), jointly with the average, over 500 replicates, of Nadaraya–Watson- and local linear-type estimators (center and right panels, respectively), using the specific scenarios considered in Figure 3.1, for model M_1 (top row panels) and model M_2 (bottom row panels).

of values of the explanatory variables. These quantities, at a point $\mathbf{x}$, are defined as:

$$\begin{aligned} \mathrm{CB}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)] &= \mathbb{E}\{\sin[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p) - m^c(\mathbf{x})]\}, && (3.35)\\ \mathrm{CVAR}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)] &= \mathbb{E}\{1-\cos[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p) - \mu(\mathbf{x};\mathbf{p})]\}, && (3.36)\\ \mathrm{CMSE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)] &= \mathbb{E}\{1-\cos[m^c(\mathbf{x}) - \hat{m}^c_{\mathbf{H}}(\mathbf{x};p)]\}, && (3.37) \end{aligned}$$

where $\mu(\mathbf{x};p)$ in CVAR denotes the circular mean of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$. Notice that, using Taylor expansions, equations (3.35), (3.36) and (3.37) are equivalent to the Euclidean versions of these expressions (Kim and SenGupta, 2017).

Figures 3.3 and 3.4 show, in the scenarios considered in Figure 3.1, the CB, CVAR and CMSE computed in a 100×100 regular grid of the explanatory variables, when using Nadaraya–Watson- and local linear-type estimators (top and bottom

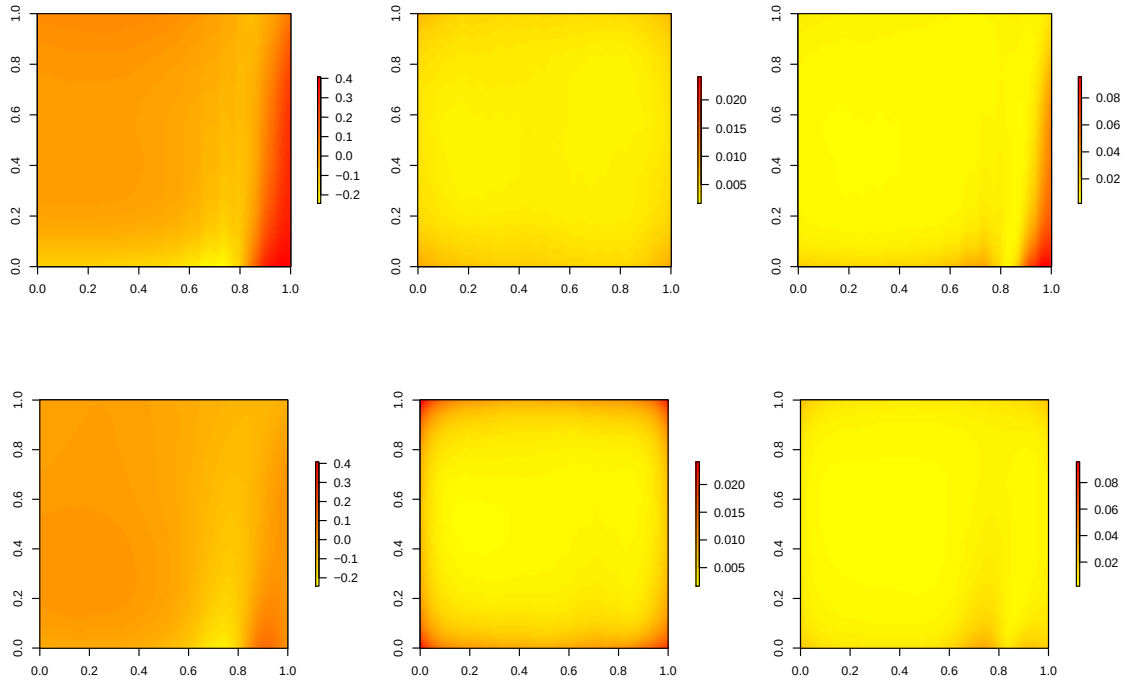

Figure 3.3: Circular bias (left panels), circular variance (center panels) and CMSE (right panels) surfaces for model M_1 for a 100×100 regular grid, using Nadaraya–Watson- and local linear-type estimators (top and bottom row panels, respectively). $n = 225$ and von Mises errors with zero mean and $\kappa = 5$.

row panels, respectively), for models M_1 and M_2, respectively. The expectations in (3.35), (3.36) and (3.37) are approximated by the averages over the 500 replicates generated. It can be seen that the Nadaraya–Watson-type estimator ($p = 0$) provides larger biases and smaller variances than the local linear-type estimator ($p = 1$) in both settings. However, the CMSE is smaller for the local linear fit in most of the grid points. Similar results for the CB, CVAR and CMSE for both estimators were obtained in other scenarios.

3.3.4 Real data illustration

A real data example is presented in order to illustrate the application of the proposed estimators. Based on the simulation study, where the local linear-type estimator

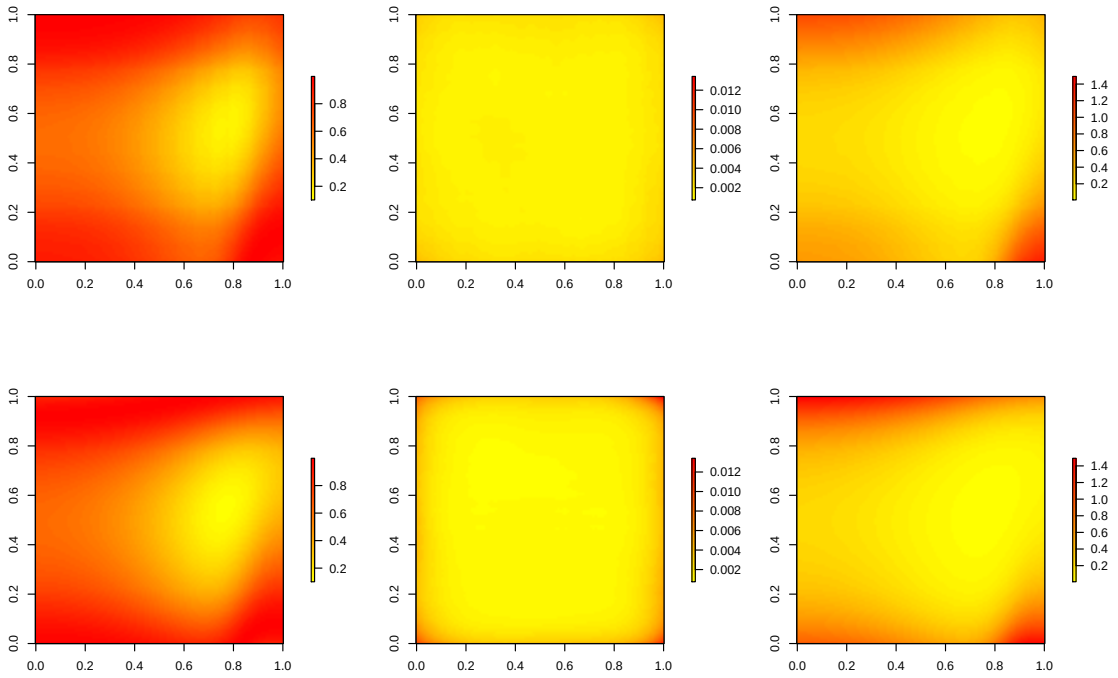

Figure 3.4: Circular bias (left panels), circular variance (center panels) and CMSE (right panles) surfaces for model M_2 for a 100×100 regular grid, using Nadaraya–Watson- and local linear-type estimators (top and bottom row panels, respectively). $n = 225$ and von Mises errors with zero mean and $\kappa = 15$.

presented a slightly better performance than the Nadaraya–Watson one, just results corresponding to $m^{c}_{\mathbf{H}}(\mathbf{x}; 1)$ are provided for real data.

Orientations of two species of sand hoppers[1] (*Talorchestia brito* and *Talitrus saltator*) on the Zouara beach in north-western Tunisia are considered. As pointed out in Scapini et al. (2002), experiments were performed in two different periods, April and October 1999. April represents the beginning of the warm season, with individuals having survived the winter, and October the end of it, with individuals

[1]The author thanks Professor Felicita Scapini and her research team who kindly provided the sand hoppers data that are used in Chapters 3 and 5 of this dissertation. Data were collected within the Project ERB ICI8-CT98-0270 from the European Commission, Directorate General XII Science.

having survived extremely warm and dry summer conditions. The sand hoppers were captured in the morning of each day with intercepting traps. The experiments were carried out using two experimental arenas, one permitting the view of both sky and landscape and the other of the sky only. A total of 777 individuals of *Talorchestia brito* and 867 *Talitrus saltator* were tested, and identified for species and sex in the laboratory.

Following the proposal in Presnell et al. (1998), these observations were analyzed in Scapini et al. (2002). They considered a projected multivariate linear model (PMLM) to analyze the orientation of two species of sand hoppers as a function of different covariates. We refer to Scapini et al. (2002) and Marchetti and Scapini (2003) for details on the experiment, a thorough data analysis and sound biological conclusions. Dealing with the same dataset, in Marchetti and Scapini (2003), the authors conclude that the orientation is different for the two sexes (males and females) and they explicitly mention that nonparametric smoothers are flexible tools that may suggest unexpected features of the data.

The illustration with our proposal is a first attempt to analyze this dataset with nonparametric tools in order to check how orientation (in degrees) behaves when temperature (in Celsius degrees) and (relative) humidity (in percentage) are included as covariates. For illustration purposes, only observations corresponding to (relative) humidity values larger than 45% are considered in this analysis. The corresponding datasets are plotted in Figure 3.5 (males in the left panel and females in the right panel), being the sample sizes $n = 330$ and $n = 404$, for male and female sand hoppers, respectively.

Figure 3.6 shows the local linear-type estimates for male (left panel) and female (right panel) mean orientations, considering temperature (horizontal axis) and relative humidity (vertical axis) as covariates. Note that configurations of temperature and humidity are the same for males and females, given that these values correspond to experimental conditions. In this example, unlike in the simulation experiments, the CV^c bandwidth matrix has been searched in the family of the symmetric and definite positive full bandwidth matrices, using an optimization algorithm based on the Nelder–Mead simplex method described in Lagarias et al. (1998). Using the

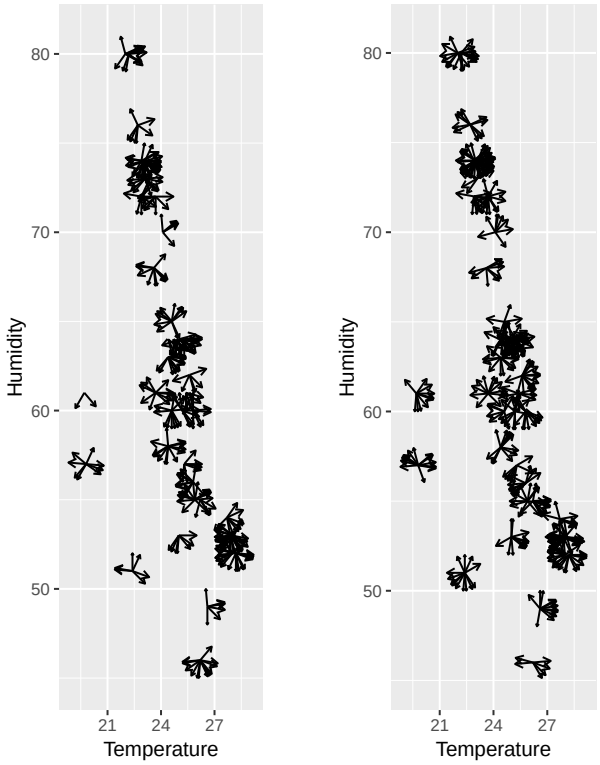

Figure 3.5: Observed orientation of male (left panel) and female (right panel) sand hoppers as a function of temperature and relative humidity.

initial bandwidth matrix $\mathbf{H}_{\text{init}} = 1.5 \cdot \text{diag}\{\hat{\sigma}_{X_1}, \hat{\sigma}_{X_2}\}$, the algorithm converged to

$$\mathbf{H}^m_{\text{CV}^c} = \begin{bmatrix} 2.7781 & 0.0001 \\ 0.0001 & 15.2529 \end{bmatrix}, \tag{3.38}$$

for males, and to

$$\mathbf{H}^f_{\text{CV}^c} = \begin{bmatrix} 4.0930 & -0.0009 \\ -0.0009 & 13.1937 \end{bmatrix}, \tag{3.39}$$

for females, where $\hat{\sigma}_{X_1}$ and $\hat{\sigma}_{X_2}$ denote the sample standard deviations of the covariates X_1 = "temperature" and X_2 = "humidity", respectively. As in the previous section, a multivariate triweight kernel is considered.

Note that the estimation grid of explanatory variables on which the estimates of the mean were computed was constructed by overlying the survey values of temper-

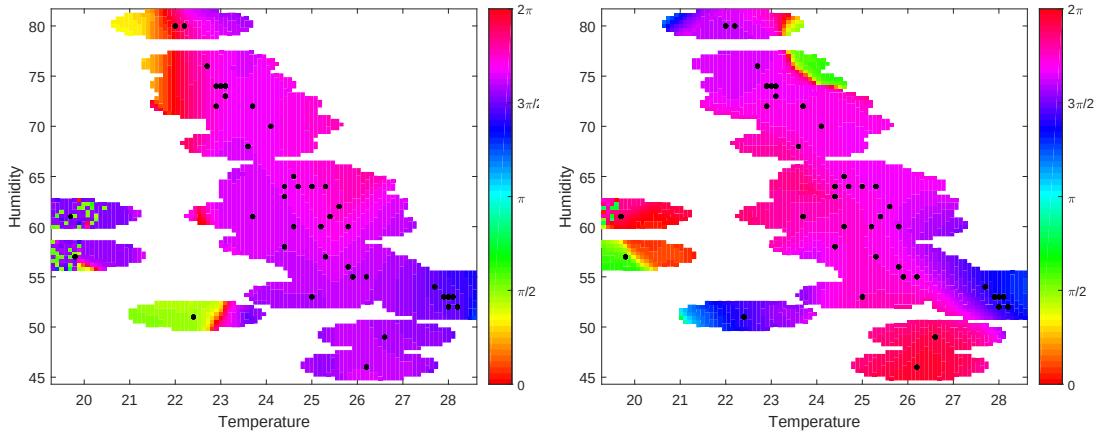

Figure 3.6: Estimates of the mean orientation of males (left panel) and females (right panel) sand hoppers, considering a local linear-type estimator with cross-validation bandwidth. Horizontal axis: temperature in Celsius degrees. Vertical axis: relative humidity in percentage.

ature and humidity with a 100×100 grid and, then, dropping every grid point that did not satisfy one of the following two requirements: (a) it is within 15 "grid cell length" from an observation point, or (b) the calculation for the estimates of the sine and cosine components at that grid point uses a smoothing vector that is sufficiently stable. Both requirements are admittedly somewhat arbitrary, but they represent a compromise between coverage over the region of interest and ability to avoid singular design matrices. Even with these restrictions, some of the estimates for low temperature values (around 20 Celsius degrees) seem to be spurious, specially in the case of male individuals. This can be due to data sparseness or a boundary effect, two well-known situations where kernel-based smoothing methods may present certain limitations. Trying to avoid some of these problems and taking into account that there are repeated values of the covariates, additional estimates have been obtained after jittering the original data (the corresponding plots are not shown), obtaining estimates that follow similar patterns to those shown in Figure 3.6. The mean direction followed by male and female sand hoppers is different for some temperature and humidity conditions. Seawards orientation was roughly $7\pi/4$, so it can be seen that females are more seawards oriented than males, specially for mid to low values of temperature.

3.4 Proofs of the main results

This section is devoted to present a detailed proof of Lemma 3.1 and Theorem 3.1, Lemma 3.2 and Theorem 3.2, Theorems 3.3 and 3.4, as well as, the proof of Lemma 3.3 and Theorem 3.5, and Lemma 3.4 and Theorem 3.6. More specifically, the asymptotic properties of the proposed nonparametric regression estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$ given in (3.14), for $p=0,1$, are established in Theorems 3.1 and 3.2, respectively. Lemmas 3.1 and 3.2 contain some previous results. For $d=1$, the extensions for $p=2$ and $p=3$ are considered in Theorems 3.3 and 3.4, respectively. The corresponding asymptotic properties of the estimators $\tilde{m}^c_{\mathbf{H}}(\mathbf{x};p)$, for $p=0,1$, given in (3.27), considering local weighted average type estimators, are provided in Theorems 3.5 and 3.6. Some previous results for local weighted average type estimators are in Lemmas 3.3 and 3.4.

For simplicity, in Taylor expansions used in the proofs of these results, $\hat{m}_{j,\mathbf{H}}$ and m_j denote $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$ and $m_j(\mathbf{x})$, respectively, for $j=1,2$. Moreover, $\hat{g}_{j,\mathbf{H}}$ and g_j stands for $\hat{g}_{j,\mathbf{H}}(\mathbf{x};p)$ and $g_j(\mathbf{x})$, respectively, for $j=1,2$.

Proof of Lemma 3.1 The asymptotic bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x};0)$, for $j=1,2$, can be directly obtained using the asymptotic properties on the multivariate Nadaraya–Watson estimator given in (1.7) and (1.8).

Regarding the conditional covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)$, using (A.1) and (A.4) of Lemma A.1, it follows that

$$\begin{aligned}\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0),\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{\sum_{i=1}^n\sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})}\\ &\quad\cdot\mathbb{C}\mathrm{ov}[\sin(\Theta_i),\cos(\Theta_j)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n]\\ &= \frac{\sum_{i=1}^n K^2_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})c(\mathbf{X}_i)}{\left[\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\right]^2}\\ &= \frac{R(K)c(\mathbf{x})}{n|\mathbf{H}|f(\mathbf{x})}+o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right),\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x}) = f(\mathbf{x})+o_{\mathbb{P}}(1)$$

and

$$\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})c(\mathbf{X}_i) = \frac{1}{|\mathbf{H}|}R(K)f(\mathbf{x})c(\mathbf{x}) + o_{\mathbb{P}}\left(|\mathbf{H}|^{-1}\right).$$

Proof of Theorem 3.1 First, to obtain the bias of the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$ given in (3.14), the function $\operatorname{atan2}(\hat{m}_{1,\mathbf{H}},\hat{m}_{2,\mathbf{H}})$ is expanded in Taylor series around (m_1,m_2), to get

$$\begin{aligned}
\operatorname{atan2}(\hat{m}_{1,\mathbf{H}},\hat{m}_{2,\mathbf{H}}) &= \operatorname{atan2}(m_1,m_2) + \frac{m_2}{m_1^2+m_2^2}(\hat{m}_{1,\mathbf{H}}-m_1)\\
&\quad -\frac{m_1}{m_1^2+m_2^2}(\hat{m}_{2,\mathbf{H}}-m_2) + \frac{m_1m_2}{(m_1^2+m_2^2)^2}(\hat{m}_{2,\mathbf{H}}-m_2)^2\\
&\quad -\frac{m_1m_2}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)^2\\
&\quad -\frac{m_1^2-m_2^2}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)(\hat{m}_{2,\mathbf{H}}-m_2)\\
&\quad +\mathcal{O}\left[(\hat{m}_{1,\mathbf{H}}-m_1)^3\right] + \mathcal{O}\left[(\hat{m}_{2,\mathbf{H}}-m_2)^3\right]. \qquad (3.40)
\end{aligned}$$

Hence, noting that $\ell(\mathbf{x}) = [m_1^2(\mathbf{x})+m_2^2(\mathbf{x})]^{1/2}$ and taking expectations in the above expression, it follows that

$$\begin{aligned}
&\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)] - m^c(\mathbf{x})\\
&\quad = \frac{m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})] - \frac{m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]\\
&\qquad +\frac{m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]^2\}\\
&\qquad -\frac{m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]^2\}\\
&\qquad -\frac{m_1^2(\mathbf{x})-m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})][\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]\}\\
&\qquad +\mathcal{O}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]^3\}.
\end{aligned}$$

Noting that $\mathbb{E}\left[(\hat{m}_{j,\mathbf{H}}-m_j)^2 \mid \mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \mathbb{V}\mathrm{ar}(\hat{m}_{j,\mathbf{H}} \mid \mathbf{X}_1,\ldots,\mathbf{X}_n) + [\mathbb{E}(\hat{m}_{j,\mathbf{H}}-$

$m_j \mid \mathbf{X}_1, \ldots, \mathbf{X}_n)]^2$, and using the results in Lemma 3.1, it is obtained that

$$\begin{aligned}\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) - m^c(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{1}{2}\frac{m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\mu_2(K)\mathrm{tr}\left[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x})\right] \\ &\quad + \frac{m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}} m_1(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla} f(\mathbf{x}) \\ &\quad - \frac{1}{2}\frac{m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\mu_2(K)\mathrm{tr}\left[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})\right] \\ &\quad - \frac{m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}} m_2(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla} f(\mathbf{x}) \\ &\quad + \boldsymbol{o}_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)].\end{aligned}$$

By straightforward calculations, one gets that

$$\begin{aligned}&\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) - m^c(\mathbf{x}) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] \\ &= \frac{1}{2}\frac{\mu_2(K)}{\ell^2(\mathbf{x})}\mathrm{tr}\left\{\mathbf{H}^2\left[m_2(\mathbf{x})\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x}) - m_1(\mathbf{x})\boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})\right]\right\} \\ &\quad + \frac{\mu_2(K)}{\ell^2(\mathbf{x})f(\mathbf{x})}\left\{\left[m_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_1(\mathbf{x}) - m_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_2(\mathbf{x})\right]\mathbf{H}^2\boldsymbol{\nabla} f(\mathbf{x})\right\} + \boldsymbol{o}_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)].\end{aligned}$$

Moreover, note that

$$\begin{aligned}\boldsymbol{\nabla} m^c(\mathbf{x}) &= \frac{1}{m_1^2(\mathbf{x}) + m_2^2(\mathbf{x})}\left[\boldsymbol{\nabla} m_1(\mathbf{x})m_2(\mathbf{x}) - \boldsymbol{\nabla} m_2(\mathbf{x})m_1(\mathbf{x})\right] \\ &= \frac{1}{\ell^2(\mathbf{x})}\left[\boldsymbol{\nabla} m_1(\mathbf{x})m_2(\mathbf{x}) - \boldsymbol{\nabla} m_2(\mathbf{x})m_1(\mathbf{x})\right], \qquad (3.41) \\ \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}) &= \frac{1}{\ell^2(\mathbf{x})}\left[\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x})m_2(\mathbf{x}) + \boldsymbol{\nabla} m_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_2(\mathbf{x})\right. \\ &\quad \left. - \boldsymbol{\nabla} m_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_1(\mathbf{x}) - \boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})m_1(\mathbf{x})\right] \\ &\quad - \frac{2}{\ell^3(\mathbf{x})}\boldsymbol{\nabla}\ell(\mathbf{x})\left[\boldsymbol{\nabla} m_1(\mathbf{x})m_2(\mathbf{x}) - \boldsymbol{\nabla} m_2(\mathbf{x})m_1(\mathbf{x})\right] \\ &= \frac{1}{\ell^2(\mathbf{x})}\left[\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x})m_2(\mathbf{x}) + \boldsymbol{\nabla} m_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_2(\mathbf{x})\right. \\ &\quad \left. - \boldsymbol{\nabla} m_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m_1(\mathbf{x}) - \boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})m_1(\mathbf{x})\right] \\ &\quad - \frac{2}{\ell(\mathbf{x})}\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x}). \qquad (3.42)\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
&\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) - m^c(\mathbf{x}) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&\quad = \frac{1}{2}\mu_2(K)\mathrm{tr}\left\{\mathbf{H}^2\left[\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}) + \frac{2}{\ell(\mathbf{x})}\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\right]\right\} \\
&\qquad + \frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x}) + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)] \\
&\quad = \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}\ell(\mathbf{x}) \\
&\qquad + \frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x}) + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)] \\
&\quad = \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)].
\end{aligned}$$

Now, in order to derive the variance of the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$, the function $\mathrm{atan2}^2(\hat{m}_{1,\mathbf{H}},\hat{m}_{2,\mathbf{H}})$ is expanded in Taylor series around (m_1,m_2), and thus, to obtain that

$$\begin{aligned}
\mathrm{atan2}^2(\hat{m}_{1,\mathbf{H}},\hat{m}_{2,\mathbf{H}}) \;=\; & \mathrm{atan2}^2(m_1,m_2) + \frac{2\mathrm{atan2}(m_1,m_2)m_2}{m_1^2+m_2^2}(\hat{m}_{1,\mathbf{H}}-m_1) \\
& - \frac{2\mathrm{atan2}(m_1,m_2)m_1}{m_1^2+m_2^2}(\hat{m}_{2,\mathbf{H}}-m_2) \\
& + \frac{2\mathrm{atan2}(m_1,m_2)m_1m_2}{(m_1^2+m_2^2)^2}(\hat{m}_{2,\mathbf{H}}-m_2)^2 \\
& - \frac{2\mathrm{atan2}(m_1,m_2)m_1m_2}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)^2 \\
& - \frac{2\mathrm{atan}(m_1,m_2)(m_1^2-m_2^2)}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)(\hat{m}_{2,\mathbf{H}}-m_2) \\
& + \frac{m_1^2}{(m_1^2+m_2^2)^2}(\hat{m}_{2,\mathbf{H}}-m_2)^2 + \frac{m_2^2}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)^2 \\
& - \frac{2m_1m_2}{(m_1^2+m_2^2)^2}(\hat{m}_{1,\mathbf{H}}-m_1)(\hat{m}_{2,\mathbf{H}}-m_2) \\
& + \mathcal{O}\left[(\hat{m}_{1,\mathbf{H}}-m_1)^3\right] + \mathcal{O}\left[(\hat{m}_{2,\mathbf{H}}-m_2)^3\right]. \qquad (3.43)
\end{aligned}$$

Taking expectations in the Taylor expansions given in (3.40) and (3.43), one gets

that

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}[\hat{m}^{c}_{\mathbf{H}}(\mathbf{x};0)] \\
&= \mathrm{atan2}^2[m_{1,\mathbf{H}}(\mathbf{x}), m_{2,\mathbf{H}}(\mathbf{x})] + \frac{2\mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})] \\
&\quad - \frac{2\mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})] \\
&\quad + \frac{2\mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]^2\} \\
&\quad - \frac{2\mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})]^2\} \\
&\quad - \frac{2\mathrm{atan}[m_1(\mathbf{x}), m_2(\mathbf{x})][m_1^2(\mathbf{x}) - m_2^2(\mathbf{x})]}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})][\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]\} \\
&\quad + \frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]^2\} + \frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})]^2\} \\
&\quad - \frac{2m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})][\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]\} \\
&\quad - \Bigg(\mathrm{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})] + \frac{m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})] \\
&\quad - \frac{m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\mathbb{E}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})] + \frac{m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]^2\} \\
&\quad - \frac{m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})]^2\} \\
&\quad - \frac{m_1^2(\mathbf{x}) - m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})][\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]\} \\
&\quad + \mathcal{O}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]^3\}\Bigg)^2 \\
&\quad + \mathcal{O}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) - m_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) - m_2(\mathbf{x})]^3\}.
\end{aligned}
$$

By straightforward calculations, it can be obtained that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)] &= \frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]^2\}+\frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]^2\}\\ &-\frac{2m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})][\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]\}\\ &-\frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\{\mathbb{E}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]\}^2-\frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}\{\mathbb{E}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]\}^2\\ &+\frac{2m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{E}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]\mathbb{E}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]\\ &+\mathcal{O}\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]^3\}+\mathcal{O}\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]^3\}.\end{aligned}$$

So, noting that $\mathbb{E}\left[(\hat{m}_{j,\mathbf{H}}-m_j)^2 \mid \mathbf{X}_1,\ldots,\mathbf{X}_n\right] = \mathbb{V}\mathrm{ar}(\hat{m}_{j,\mathbf{H}} \mid \mathbf{X}_1,\ldots,\mathbf{X}_n)+[\mathbb{E}(\hat{m}_{j,\mathbf{H}}-m_j \mid \mathbf{X}_1,\ldots,\mathbf{X}_n)]^2$, it can be obtained that the conditional variance is:

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n]\\ &+\frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n]\\ &-\frac{2m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0),\hat{m}_{2,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n]\\ &+\mathcal{O}\left\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)-m_1(\mathbf{x})]^3\right\}+\mathcal{O}\left\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)-m_2(\mathbf{x})]^3\right\}.\end{aligned}$$

Therefore, using Lemma 3.1, one gets that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{n|\mathbf{H}|}R(K)\frac{m_1^2(\mathbf{x})s_2^2(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})}+\frac{1}{n|\mathbf{H}|}R(K)\frac{m_2^2(\mathbf{x})s_1^2(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})}\\ &-\frac{2}{n|\mathbf{H}|}R(K)\frac{m_1(\mathbf{x})m_2(\mathbf{x})c(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})}+o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

Considering (3.8), (3.11), (3.12) and (3.13), and taking into account that $f_1^2(\mathbf{x})+$

$f_2^2(\mathbf{x}) = 1$, it follows that

$$\begin{aligned}
&m_1^2(\mathbf{x})s_2^2(\mathbf{x}) + m_2^2(\mathbf{x})s_1^2(\mathbf{x}) - 2m_1(\mathbf{x})m_2(\mathbf{x})c(\mathbf{x}) \\
&= f_1^2(\mathbf{x})f_2^2(\mathbf{x})\ell^2(\mathbf{x})\sigma_2^2(\mathbf{x}) - 2f_2(\mathbf{x})f_1^3(\mathbf{x})\ell^2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_1^4(\mathbf{x})\ell^2(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&\quad + f_2^2(\mathbf{x})f_1^2(\mathbf{x})\ell^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_2^3(\mathbf{x})f_1(\mathbf{x})\ell^2(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^4(\mathbf{x})\ell^2(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&\quad - 2f_1^2(\mathbf{x})f_2^2(\mathbf{x})\ell^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_1^3(\mathbf{x})f_2(\mathbf{x})\ell^2(\mathbf{x})\sigma_{12}(\mathbf{x}) - 2f_2^3(\mathbf{x})f_1(\mathbf{x})\ell^2(\mathbf{x})\sigma_{12}(\mathbf{x}) \\
&\quad + 2f_1^2(\mathbf{x})f_2^2(\mathbf{x})\ell^2(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&= \ell^2(\mathbf{x})\sigma_1^2(\mathbf{x}). \qquad (3.44)
\end{aligned}$$

Therefore,

$$\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Proof of Lemma 3.2 The asymptotic bias and variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x};1)$, for $j = 1, 2$, can be derived using the asymptotic properties on the multivariate local linear estimator given in (1.10) and (1.11). In this case, the conditional covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};1)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};1)$ is:

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1), \hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&= \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}^c\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\mathbf{e}_1,
\end{aligned}$$

where $\boldsymbol{\Sigma}^c$ is the covariance matrix of $\sin(\Theta)$ and $\cos(\Theta)$, whose (i,j)-entry is $\boldsymbol{\Sigma}^c(i,j) = \mathbb{C}\mathrm{ov}[\sin(\Theta_i), \cos(\Theta_j) \mid \mathbf{X}_i, \mathbf{X}_j]$, $i,j = 1,\ldots,n$. Notice that $\boldsymbol{\Sigma}^c(i,j) = 0$, for $i \neq j$. Defining

$$\begin{aligned}
k_{1,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}), \\
k_{2,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x}), \\
k_{3,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathrm{T}}, \\
s_{31,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i=1}^n K^2_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})c(\mathbf{X}_i), \\
s_{32,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i=1}^n K^2_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})c(\mathbf{X}_i), \\
s_{33,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i=1}^n K^2_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})(\mathbf{X}_i - \mathbf{x})^{\mathrm{T}}c(\mathbf{X}_i),
\end{aligned}$$

and using (A.1), (A.2), (A.3) (A.4), (A.5) and (A.6) of Lemma A.1, it follows that

$$\begin{aligned}
k_{1,n}(\mathbf{x}) &= f(\mathbf{x}) + o_{\mathbb{P}}(1),\\
k_{2,n}(\mathbf{x}) &= \mu_2(K)\boldsymbol{\nabla} f(\mathbf{x})\mathbf{H}^2 + o_{\mathbb{P}}(\mathbf{H}^2\mathbf{1}_d),\\
k_{3,n}(\mathbf{x}) &= \mu_2(K) f(\mathbf{x})\mathbf{H}^2 + o_{\mathbb{P}}(\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H}),\\
s_{31,n}(\mathbf{x}) &= \frac{1}{n|\mathbf{H}|} R(K) f(\mathbf{x}) c(\mathbf{x}) + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \qquad (3.45)\\
s_{32,n}(\mathbf{x}) &= \frac{1}{n|\mathbf{H}|} o_{\mathbb{P}}(\mathbf{1}_d), \qquad (3.46)\\
s_{33,n}(\mathbf{x}) &= \frac{1}{n|\mathbf{H}|} o_{\mathbb{P}}(\mathbf{1}_{d\times d}). \qquad (3.47)
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left(n^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}}\right)^{-1} &= \begin{pmatrix} k_{1,n}(\mathbf{x}) & k_{2,n}^{\mathrm{T}}(\mathbf{x}) \\ k_{2,n}(\mathbf{x}) & k_{3,n}(\mathbf{x}) \end{pmatrix}^{-1}\\
&= \begin{pmatrix} \frac{1}{f(\mathbf{x})} + o_{\mathbb{P}}(1) & \frac{-\boldsymbol{\nabla}^{\mathrm{T}} f(\mathbf{x})}{f^2(\mathbf{x})} + o_{\mathbb{P}}(\mathbf{1}_d^{\mathrm{T}}) \\ \frac{-\boldsymbol{\nabla} f(\mathbf{x})}{f^2(\mathbf{x})} + o_{\mathbb{P}}(\mathbf{1}_d) & \frac{1}{\mu_2(K) f(\mathbf{x})\mathbf{H}^2} + o_{\mathbb{P}}(\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H}) \end{pmatrix}, \qquad (3.48)
\end{aligned}$$

$$\begin{aligned}
n^{-2}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}^{c}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}} &= \begin{pmatrix} s_{31,n}(\mathbf{x}) & s_{32,n}^{\mathrm{T}}(\mathbf{x}) \\ s_{32,n}(\mathbf{x}) & s_{33,n}(\mathbf{x}) \end{pmatrix}\\
&= \frac{1}{n|\mathbf{H}|}\begin{pmatrix} R(K) f(\mathbf{x}) c(\mathbf{x}) + o_{\mathbb{P}}(1) & o_{\mathbb{P}}(\mathbf{1}_d^{\mathrm{T}}) \\ o_{\mathbb{P}}(\mathbf{1}_d) & o_{\mathbb{P}}(\mathbf{1}_{d\times d}) \end{pmatrix}.
\end{aligned}$$

Consequently, by straightforward calculations, one gets

$$\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1), \hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n|\mathbf{H}| f(\mathbf{x})} R(K) c(\mathbf{x}) + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Proof of Theorem 3.2 To obtain the bias of the estimator $\hat{m}^{c}_{\mathbf{H}}(\mathbf{x};1)$ given in (3.14), following the arguments used in the proof of Theorem 3.1 and using results

in Lemma 3.2, one gets that

$$\begin{aligned}\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1) - m^c(\mathbf{x}) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{2}\mu_2(K)\frac{m_2(\mathbf{x})}{\ell^2(\mathbf{x})}\mathrm{tr}\left[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x})\right] \\ &\quad -\frac{1}{2}\mu_2(K)\frac{m_1(\mathbf{x})}{\ell^2(\mathbf{x})}\mathrm{tr}\left[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})\right] + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)] \\ &= \frac{1}{2}\frac{\mu_2(K)}{\ell^2(\mathbf{x})}\mathrm{tr}\left\{\mathbf{H}^2\left[m_2(\mathbf{x})\boldsymbol{\mathcal{H}}_{m_1}(\mathbf{x}) - m_1(\mathbf{x})\boldsymbol{\mathcal{H}}_{m_2}(\mathbf{x})\right]\right\} \\ &\quad + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)].\end{aligned}$$

Considering (3.41) and (3.42), it can be obtained that

$$\begin{aligned}&\mathbb{E}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1) - m^c(\mathbf{x}) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\ &= \frac{1}{2}\mu_2(K)\mathrm{tr}\left\{\mathbf{H}^2\left[\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}) + \frac{2}{\ell(x)}\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\right]\right\} + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)] \\ &= \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}\ell(\mathbf{x}) + o_{\mathbb{P}}[\mathrm{tr}(\mathbf{H}^2)].\end{aligned}$$

As for the variance of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$, using (3.44) and Lemma 3.2, it follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{n|\mathbf{H}|}R(K)\frac{m_1^2(\mathbf{x})s_2^2(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})} + \frac{1}{n|\mathbf{H}|}R(K)\frac{m_2^2(\mathbf{x})s_1^2(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})} \\ &\quad -\frac{2}{n|\mathbf{H}|}R(K)\frac{m_1(\mathbf{x})m_2(\mathbf{x})c(\mathbf{x})}{\ell^4(\mathbf{x})f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right) \\ &= \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

.

Proof of Theorem 3.3 Using the asymptotic properties of the local quadratic estimator given in (1.14) and (1.16), close expressions of $\mathbb{E}[\hat{m}_{j,h}(x;2) \mid X_1,\ldots,X_n]$ and $\mathbb{V}\mathrm{ar}[\hat{m}_{j,h}(x;2) \mid X_1,\ldots,X_n]$, for $j = 1,2$, can be obtained. To derive the bias of $\hat{m}^c_h(x;2)$, following the arguments used in the proof of Theorems 3.1 and 3.2, one

gets that

$$\begin{aligned}\mathbb{E}[\hat{m}_h^c(x;2) - m^c(x) \mid X_1,\ldots,X_n] &= \frac{h^4\mu_4(K_{(2)})f^{(1)}(x)}{3!f(x)}\frac{m_2(x)}{\ell^2(x)}m_1^{(3)}(x)\\ &+\frac{h^4\mu_4(K_{(2)})}{4!}\frac{m_2(x)}{\ell^2(x)}m_1^{(4)}(x)\\ &-\frac{h^4\mu_4(K_{(2)})f^{(1)}(x)}{3!f(x)}\frac{m_1(x)}{\ell^2(x)}m_2^{(3)}(x)\\ &-\frac{h^4\mu_4(K_{(2)})}{4!}\frac{m_1(x)}{m_1^2(x)+m_1^2(x)}m_2^{(4)}(x)+o_{\mathbb{P}}(h^4).\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[\hat{m}_h^c(x;2) - m^c(x) \mid X_1,\ldots,X_n] &= \frac{h^4\mu_4(K_{(2)})f^{(1)}(x)}{3!f(x)\ell^2(x)}[m_2(x)m_1^{(3)}(x) - m_1(x)m_2^{(3)}(x)]\\ &+\frac{h^4\mu_4(K_{(2)})}{4!\ell^2(x)}[m_2(x)m_1^{(4)}(x) - m_1(x)m_2^{(4)}(x)]\\ &+o_{\mathbb{P}}(h^4).\end{aligned}$$

Now, taking into account that

$$\begin{aligned}m^{c(1)}(x) &= \frac{1}{\ell^2(x)}\left[m_1^{(1)}(x)m_2(x) - m_2^{(1)}(x)m_1(x)\right],\\ m^{c(2)}(x) &= \frac{1}{\ell^2(x)}\left[m_1^{(2)}(x)m_2(x) - m_2^{(2)}(x)m_1(x)\right] - \frac{2}{\ell(x)}\ell^{(1)}(x)m^{c(1)}(x),\\ m^{c(3)}(x) &= \frac{1}{\ell^2(x)}\left[m_1^{(3)}(x)m_2(x) - m_2^{(3)}(x)m_1(x) + m_1^{(2)}(x)m_2^{(1)}(x) - m_1^{(1)}(x)m_2^{(2)}(x)\right]\\ &-\frac{4}{\ell(x)}\ell^{(1)}(x)m^{c(2)}(x) - \frac{2}{\ell^2(x)}{\ell^{(1)}}^2(x)m^{c(1)}(x) - \frac{2}{\ell(x)}\ell^{(2)}(x)m^{c(1)}(x),\\ m^{c(4)}(x) &= \frac{1}{\ell^2(x)}\left[m_1^{(4)}(x)m_2(x) - m_2^{(4)}(x)m_1(x) + 2m_1^{(3)}(x)m_2^{(1)}(x) - 2m_1^{(1)}(x)m_2^{(3)}(x)\right]\\ &-\frac{6}{\ell(x)}\ell^{(1)}(x)m^{(3)}(x) - \frac{2}{\ell(x)}\ell^{(3)}(x)m^{c(1)}(x) - \frac{6}{\ell(x)}\ell^{(2)}(x)m^{c(2)}(x)\\ &-\frac{6}{\ell(x)^2}{\ell^{(1)}}^2(x)m^{c(2)}(x) - \frac{6}{\ell(x)^2}\ell^{(1)}(x)\ell^{(2)}(x)m^{c(1)}(x),\end{aligned}$$

it follows that

$$\begin{aligned}
&\mathbb{E}[\hat{m}_h^c(x;2) - m^c(x) \mid X_1,\ldots,X_n] \\
&= \frac{h^4\mu_4(K_{(2)})f^{(1)}(x)}{3!f(x)}\left[m^{(3)}(x) + \frac{2\ell^{(2)}(x)m^{c(1)}(x)}{\ell(x)} + \frac{m_2^{(2)}(x)m_1^{(1)}(x) - m_1^{(2)}(x)m_2^{(1)}(x)]}{\ell^2(x)}\right. \\
&\quad + \left.\frac{4\ell^{(1)}(x)m^{c(2)}(x)}{\ell(x)} + \frac{2{\ell^{(1)}}^2(x)m^{c(1)}(x)}{\ell^2(x)}\right] + \frac{h^4\mu_4(K_{(2)})}{4!}\left[m^{(4)}(x) + \frac{2\ell^{(3)}(x)m^{c(1)}(x)}{\ell(x)}\right. \\
&\quad + \frac{2m_2^{(3)}(x)m_1^{(1)}(x) - 2m_1^{(3)}(x)m_2^{(1)}(x)}{\ell^2(x)} + \frac{6\ell^{(1)}(x)m^{(3)}(x) + 6\ell^{(2)}(x)m^{c(2)}(x)}{\ell(x)} \\
&\quad + \left.\frac{6{\ell^{(1)}}^2(x)m^{c(2)}(x) + 6\ell^{(1)}(x)\ell^{(2)}(x)m^{c(1)}(x)}{\ell^2(x)}\right] + o_{\mathbb{P}}\left(h^4\right).
\end{aligned}$$

As for the variance of $\hat{m}_h^c(x;2)$, the same arguments as those employed in the proof of Theorem 3.1 and 3.2 can be used. The conditional covariance between both $\hat{m}_{1,h}(x;2)$ and $\hat{m}_{2,h}(x;2)$ is

$$\mathbb{C}\mathrm{ov}[\hat{m}_{1,h}(x;2), \hat{m}_{2,h}(x;2) \mid X_1,\ldots,X_n] = \frac{1}{nhf(x)}R(K_{(2)})c(x) + o_{\mathbb{P}}\left(\frac{1}{nh}\right),$$

and, therefore, the variance of $\hat{m}_h^c(x;2)$ is:

$$\mathbb{V}\mathrm{ar}[\hat{m}_h^c(x;2) \mid X_1,\ldots,X_n] = \frac{1}{nh\ell^2(x)f(x)}R(K_{(2)})\sigma_1^2(x) + o_{\mathbb{P}}\left(\frac{1}{nh}\right).$$

Proof of Theorem 3.4 To obtain the conditional bias of $\hat{m}_h^c(x;3)$, using the asymptotic properties of the local cubic estimator given in (1.15) and (1.16), one gets that

$$\begin{aligned}
&\mathbb{E}[\hat{m}_h^c(x;3) - m^c(x) \mid X_1,\ldots,X_n] \\
&= \frac{h^4\mu_4(K_{(2)})}{4!}\frac{m_2(x)}{\ell^2(x)}m_1^{(4)}(x) - \frac{h^4\mu_4(K_{(2)})}{4!}\frac{m_1(x)}{m_1^2(x)+m_1^2(x)}m_2^{(4)}(x) + o_{\mathbb{P}}(h^2) \\
&= \frac{h^4\mu_4(K_{(2)})}{4!}\left[m^{(4)}(x) + \frac{2\ell^{(3)}(x)m^{c(1)}(x)}{\ell(x)} + \frac{2m_2^{(3)}(x)m_1^{(1)}(x) - 2m_1^{(3)}(x)m_2^{(1)}(x)}{\ell^2(x)}\right. \\
&\quad + \left.\frac{6\ell^{(1)}(x)m^{(3)}(x) + 6\ell^{(2)}(x)m^{c(2)}(x)}{\ell(x)} + \frac{6{\ell^{(1)}}^2(x)m^{c(2)}(x) + 6\ell^{(1)}(x)\ell^{(2)}(x)m^{c(1)}(x)}{\ell^2(x)}\right] \\
&\quad + o_{\mathbb{P}}\left(h^4\right).
\end{aligned}$$

Reasoning as in the proof of Theorem 3.4, the conditional variance of $\hat{m}_h^c(x;3)$ can be obtained:

$$\mathbb{V}\mathrm{ar}[\hat{m}_h^c(x;3) \mid X_1,\ldots,X_n] = \frac{1}{nh\ell^2(x)f(x)} R(K_{(2)})\sigma_1^2(x) + o_{\mathbb{P}}\left(\frac{1}{nh}\right).$$

Proof of Lemma 3.3 First, using the result for the conditional mean $\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] = \mathbb{E}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}$, and that

$$\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n}\sum_{i=1}^{n} m_j(\mathbf{X}_i)\hat{W}_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x};0), \tag{3.49}$$

with $\hat{W}_{\mathbf{H}}(\mathbf{u};0)$ given in (3.30), under assumptions (A1), (C1), (H1) and (K1), it follows that,

$$\begin{aligned}
\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] &= \int m_j(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u} = \int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x})m_j(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \int K(\mathbf{p})m_j(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= \int K(\mathbf{p})[m_j(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}] \\
&\quad \cdot[f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}]d\mathbf{p} + o[\mathrm{tr}(\mathbf{H}^2)] \\
&= f(\mathbf{x})m_j(\mathbf{x})\int K(\mathbf{p})d\mathbf{p} + \int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x})d\mathbf{p} \\
&\quad + \frac{1}{2}m_j(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} \\
&\quad + \frac{1}{2}f(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} + o[\mathrm{tr}(\mathbf{H}^2)].
\end{aligned}$$

By straightforward calculations, it can be proved that

$$\begin{aligned}
\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} &= \mathrm{tr}\left[\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\int K(\mathbf{p})\mathbf{p}\mathbf{p}^{\mathrm{T}}d\mathbf{p}\right] \\
&= \mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_f(\mathbf{x})],
\end{aligned} \tag{3.50}$$

and, similarly,

$$\begin{aligned}\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x})d\mathbf{p} &= \mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\nabla}m_j(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})],\\ \int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} &= \mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})].\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] &= f(\mathbf{x})m_j(\mathbf{x}) + \mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\nabla}m_j(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})]\\ &\quad +\frac{1}{2}\mu_2(K)m_j(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_f(\mathbf{x})]\\ &\quad +\frac{1}{2}\mu_2(K)f(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})] + o[\mathrm{tr}(\mathbf{H}^2)]\\ &= g_j(\mathbf{x}) + \frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x})] + o[\mathrm{tr}(\mathbf{H}^2)],\end{aligned}$$

because of $g_j(\mathbf{x}) = m_j(\mathbf{x})f(\mathbf{x}), j = 1, 2$, and

$$\boldsymbol{\nabla}g_j(\mathbf{x}) = \boldsymbol{\nabla}m_j(\mathbf{x})f(\mathbf{x}) + \boldsymbol{\nabla}f(\mathbf{x})m_j(\mathbf{x}), \tag{3.51}$$

$$\begin{aligned}\boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x}) &= \boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})f(\mathbf{x}) + \boldsymbol{\nabla}m_j(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x}) + \boldsymbol{\nabla}f(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m_j(\mathbf{x})\\ &\quad +\boldsymbol{\mathcal{H}}_f(\mathbf{x})m_j(\mathbf{x}).\end{aligned} \tag{3.52}$$

Concerning the variance of $\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)$, for $j = 1, 2$, it follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] &= \mathbb{E}\{\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\ &\quad +\mathbb{V}\mathrm{ar}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}.\end{aligned}$$

Firstly,

$$\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n^2}\sum_{i=1}^{n} s_j^2(\mathbf{X}_i)\hat{W}_{\mathbf{H}}^2(\mathbf{X}_i - \mathbf{x};0). \tag{3.53}$$

Therefore, considering (3.30) and under assumptions (A1), (C1), (H1) and (K1),

$$\begin{aligned}\mathbb{E}\{\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\} &= \frac{1}{n}\int s_j^2(\mathbf{u})\hat{W}_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u}\\ &= \frac{1}{n}\int K_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x})s_j^2(\mathbf{u})f(\mathbf{u})\\ &= \frac{1}{n|\mathbf{H}|}\int K^2(\mathbf{p})s_j^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}\\ &= \frac{1}{n|\mathbf{H}|}\int K^2(\mathbf{p})[s_j^2(\mathbf{x})+o(1)][f(\mathbf{x})+o(1)]d\mathbf{p}\\ &= \frac{1}{n|\mathbf{H}|}R(K)s_j^2(\mathbf{x})f(\mathbf{x})+o\left(\frac{1}{n|\mathbf{H}|}\right). \qquad (3.54)\end{aligned}$$

On the other hand, considering (3.30) and (3.49), and under assumptions (A1), (C1), (H1) and (K1), it follows that,

$$\begin{aligned}&\mathbb{V}\mathrm{ar}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\ &= \frac{1}{n}\int m_j^2(\mathbf{u})\hat{W}_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u}\\ &\quad -\frac{1}{n}\int m_j(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u}\int m_j(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};0)f(\mathbf{v})d\mathbf{v}\\ &= \frac{1}{n|\mathbf{H}|}R(K)m_j^2(\mathbf{x})f(\mathbf{x})-\frac{1}{n}g_j^2(\mathbf{x})+o\left(\frac{1}{n|\mathbf{H}|}\right). \qquad (3.55)\end{aligned}$$

Hence, from (3.53), (3.54), (3.55), and observing that $\nu_j^2(\mathbf{x}) = m_j^2(\mathbf{x})+s_j^2(\mathbf{x})$, one gets that

$$\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};0)] = \frac{1}{n|\mathbf{H}|}R(K)\nu_j^2(\mathbf{x})f(\mathbf{x})+o\left(\frac{1}{n|\mathbf{H}|}\right).$$

Finally, for the covariance between $\hat{g}_{1,\mathbf{H}}(\mathbf{x};0)$ and $\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)$, taking into account that

$$\begin{aligned}&\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0),\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)]\\ &= \mathbb{E}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0)\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\ &\quad -\mathbb{E}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\mathbb{E}\{\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\},\end{aligned}$$

and that

$$\begin{aligned}
&\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0)\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)|\mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&\quad = \frac{1}{n^2}\sum_{i=1}^{n}\tilde{c}(\mathbf{X}_i)\hat{W}^2_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x};0) \\
&\qquad +\frac{1}{n^2}\sum_{i\neq j}^{n} m_1(\mathbf{X}_i)m_2(\mathbf{X}_j)\hat{W}_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x};0)\hat{W}_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x};0),
\end{aligned}$$

under assumptions (A1), (C1), (H1) and (K1), from (3.30) it follows that

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0),\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)] \\
&\quad = \frac{1}{n}\int \tilde{c}(\mathbf{u})\hat{W}^2_{\mathbf{H}}(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u} \\
&\qquad +\frac{n-1}{n}\int m_1(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u}\int m_2(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};0)f(\mathbf{v})d\mathbf{v} \\
&\qquad -\int m_1(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};0)f(\mathbf{u})d\mathbf{u}\int m_2(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};0)f(\mathbf{v})d\mathbf{v} \\
&\quad = \frac{1}{n|\mathbf{H}|}R(K)\tilde{c}(\mathbf{x})f(\mathbf{x})+o\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

Proof of Theorem 3.5 First, to obtain the bias of the estimator $\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)$ given in (3.27), using the same linearization arguments as in the proof of Theorem 3.1, the function $\mathrm{atan2}(\hat{g}_{1,\mathbf{H}},\hat{g}_{2,\mathbf{H}})$ is expanded in Taylor series around (g_1,g_2):

$$\begin{aligned}
\mathrm{atan2}(\hat{g}_{1,\mathbf{H}},\hat{g}_{2,\mathbf{H}}) &= \mathrm{atan2}(g_1,g_2)+\frac{f_2}{\ell f}(\hat{g}_{1,\mathbf{H}}-g_1)-\frac{f_1}{\ell f}(\hat{g}_{2,\mathbf{H}}-g_2) \\
&\quad +\frac{f_1f_2}{\ell^2 f^2}(\hat{g}_{2,\mathbf{H}}-g_2)^2-\frac{f_1f_2}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}}-g_1)^2 \\
&\quad -\frac{f_1^2-f_2^2}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}}-g_1)(\hat{g}_{2,\mathbf{H}}-g_2) \\
&\quad +\mathcal{O}[(\hat{g}_{1,\mathbf{H}}-g_1)^3]+\mathcal{O}[(\hat{g}_{2,\mathbf{H}}-g_2)^3]. \qquad (3.56)
\end{aligned}$$

Hence, taking expectations in the above expression, it follows that

$$\begin{aligned}\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] - m^c(\mathbf{x}) &= \frac{f_2(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})] \\ &\quad -\frac{f_1(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})] \\ &\quad +\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^2\} \\ &\quad -\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^2\} \\ &\quad -\frac{f_1^2(\mathbf{x}) - f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]\} \\ &\quad +\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^3\}.\end{aligned}$$

Noting that $\mathbb{E}[(\hat{g}_i - g_i)^2] = \mathbb{V}\mathrm{ar}(\hat{g}_i) + [\mathbb{E}(\hat{g}_i) - g_i]^2$, and using the results in Lemma 3.3, one gets that

$$\begin{aligned}\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] - m^c(\mathbf{x}) &= \frac{f_2(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_1}(\mathbf{x})] \\ &\quad -\frac{f_1(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_2}(\mathbf{x})] + \mathcal{O}\left(\frac{1}{n|\mathbf{H}|}\right) + o[\mathrm{tr}(\mathbf{H}^2)].\end{aligned}$$

Now, recalling that $g_j(\mathbf{x}) = \ell(\mathbf{x})f(\mathbf{x})f_j(\mathbf{x}), j = 1, 2$, it follows that

$$\begin{aligned}\boldsymbol{\nabla} g_j(\mathbf{x}) &= \boldsymbol{\nabla}(\ell f)(\mathbf{x})f_j(\mathbf{x}) + \boldsymbol{\nabla} f_j(\mathbf{x})(\ell f)(\mathbf{x}), \\ \boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x}) &= \boldsymbol{\mathcal{H}}_{(\ell f)}(\mathbf{x})f_j(\mathbf{x}) + \boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_j(\mathbf{x}) \\ &\quad +\boldsymbol{\nabla} f_j(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x}) + \boldsymbol{\mathcal{H}}_{f_j}(\mathbf{x})(\ell f)(\mathbf{x}).\end{aligned}$$

Therefore,

$$\begin{aligned}&f_2(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_1}(\mathbf{x})] - f_1(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_2}(\mathbf{x})] \\ &= \mathrm{tr}\{\mathbf{H}^2[f_2(\mathbf{x})\boldsymbol{\mathcal{H}}_{(\ell f)}(\mathbf{x})f_1(\mathbf{x}) + f_2(\mathbf{x})\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_1(\mathbf{x}) + f_2(\mathbf{x})\boldsymbol{\nabla} f_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x}) \\ &\quad +f_2(\mathbf{x})\boldsymbol{\mathcal{H}}_{f_2}(\mathbf{x})(\ell f)(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\mathcal{H}}_{(\ell f)}(\mathbf{x})f_2(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_2(\mathbf{x}) \\ &\quad -f_1(\mathbf{x})\boldsymbol{\nabla} f_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\mathcal{H}}_{f_2}(\mathbf{x})(\ell f)(\mathbf{x})]\}\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}(\mathbf{H}^2\{\boldsymbol{\nabla}(\ell f)(\mathbf{x})[f_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_1(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_2(\mathbf{x})] \\
&\quad +[f_2(\mathbf{x})\boldsymbol{\nabla} f_1(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla} f_2(\mathbf{x})]\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x}) \\
&\quad +(\ell f)(\mathbf{x})[f_2(\mathbf{x})\boldsymbol{\mathcal{H}}_{f_1}(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\mathcal{H}}_{f_2}(\mathbf{x})]\}).
\end{aligned}
$$

Taking into account that $m^c(\mathbf{x}) = \operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})]$, it can be obtained that

$$
\begin{aligned}
\boldsymbol{\nabla} m^c(\mathbf{x}) &= \boldsymbol{\nabla} f_1(\mathbf{x}) f_2(\mathbf{x}) - \boldsymbol{\nabla} f_2(\mathbf{x}) f_1(\mathbf{x}), && (3.57) \\
\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}) &= \boldsymbol{\mathcal{H}}_{f_1}(\mathbf{x}) f_2(\mathbf{x}) + \boldsymbol{\nabla} f_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_1(\mathbf{x}) - \boldsymbol{\nabla} f_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_2(\mathbf{x}) \\
&\quad -\boldsymbol{\mathcal{H}}_{f_2}(\mathbf{x}) f_1(\mathbf{x}). && (3.58)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&f_2(\mathbf{x})\operatorname{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_1}(\mathbf{x})] - f_1(\mathbf{x})\operatorname{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_2}(\mathbf{x})] \\
&\quad = \operatorname{tr}(\mathbf{H}^2\{\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x}) + \boldsymbol{\nabla} m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x}) \\
&\qquad +(\ell f)(\mathbf{x})[\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x}) - \boldsymbol{\nabla} f_2(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_1(\mathbf{x}) + \boldsymbol{\nabla} f_1(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} f_2(\mathbf{x})]\}) \\
&\quad = \operatorname{tr}\{\mathbf{H}^2[2\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x}) + \ell(\mathbf{x}) f(\mathbf{x})\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})]\} \\
&\quad = 2\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) + \operatorname{tr}[\ell(\mathbf{x}) f(\mathbf{x})\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})]. \qquad (3.59)
\end{aligned}
$$

Then,

$$
\begin{aligned}
&\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] - m^c(\mathbf{x}) \\
&\quad = \frac{\mu_2(K)}{\ell(\mathbf{x}) f(\mathbf{x})}\frac{1}{2}\{2\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) + \operatorname{tr}[\ell(\mathbf{x}) f(\mathbf{x})\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})]\} + o[\operatorname{tr}(\mathbf{H}^2)] \\
&\quad = \frac{1}{2}\mu_2(K)\operatorname{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x}) f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}} m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) + o[\operatorname{tr}(\mathbf{H}^2)].
\end{aligned}
$$

To derive the variance, the function $\operatorname{atan2}^2(\hat{g}_{1,\mathbf{H}}, \hat{g}_{2,\mathbf{H}})$ is expanded in Taylor series around (g_1, g_2), as follows:

$$
\begin{aligned}
\operatorname{atan2}^2(\hat{g}_{1,\mathbf{H}}, \hat{g}_{2,\mathbf{H}}) &= \operatorname{atan2}^2(g_{1,\mathbf{H}}, g_{2,\mathbf{H}}) + \frac{2\operatorname{atan2}(g_1, g_2) f_2}{\ell f}(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}}) \\
&\quad - \frac{2\operatorname{atan2}(g_1, g_2) f_1}{\ell f}(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}}) \\
&\quad + \frac{2\operatorname{atan2}(g_1, g_2) f_1 f_2}{\ell^2 f^2}(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}})^2 \\
&\quad - \frac{2\operatorname{atan2}(g_1, g_2) f_1 f_2}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}})^2 \\
&\quad - \frac{2\operatorname{atan}(g_1, g_2)(f_1^2 - f_2^2)}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}})(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}}) \\
&\quad + \frac{f_1^2}{\ell^2 f^2}(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}})^2 + \frac{f_2^2}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}})^2 \\
&\quad - \frac{2 f_1 f_2}{\ell^2 f^2}(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}})(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}}) \\
&\quad + \mathcal{O}[(\hat{g}_{1,\mathbf{H}} - g_{1,\mathbf{H}})^3] + \mathcal{O}[(\hat{g}_{2,\mathbf{H}} - g_{2,\mathbf{H}})^3]. \qquad (3.60)
\end{aligned}
$$

Now, noting that $\mathbb{V}\mathrm{ar}(\tilde{m}^c_{\mathbf{H}}) = \mathbb{E}\left[(\tilde{m}^c_{\mathbf{H}})^2\right] - [\mathbb{E}(\tilde{m}^c_{\mathbf{H}})]^2$, and taking expectations in the expressions (3.56) and (3.60), it can be obtained that

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x}; 0)] \\
&= \operatorname{atan2}^2[g_{1,\mathbf{H}}(\mathbf{x}), g_{2,\mathbf{H}}(\mathbf{x})] + \frac{2\operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_2(\mathbf{x})}{\ell(\mathbf{x}) f(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 0) - g_1(\mathbf{x})] \\
&\quad - \frac{2\operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x})}{\ell(\mathbf{x}) f(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x}; 0) - g_2(\mathbf{x})] \\
&\quad + \frac{2\operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x}) f_2(\mathbf{x})}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x}; 0) - g_2(\mathbf{x})]^2\} \\
&\quad - \frac{2\operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x}) f_2(\mathbf{x})}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 0) - g_1(\mathbf{x})]^2\} \\
&\quad - \frac{2\operatorname{atan}[g_1(\mathbf{x}), g_2(\mathbf{x})][f_1^2(\mathbf{x}) - f_2^2(\mathbf{x})]}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 0) - g_1(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x}; 0) - g_2(\mathbf{x})]\} \\
&\quad + \frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x}; 0) - g_2(\mathbf{x})]^2\} + \frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 0) - g_1(\mathbf{x})]^2\} \\
&\quad - \frac{2 f_1(\mathbf{x}) f_2(\mathbf{x})}{\ell^2(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}; 0) - g_1(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x}; 0) - g_2(\mathbf{x})]\}.
\end{aligned}
$$

$$
\begin{aligned}
&-\Bigg(\operatorname{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] + \frac{f_2(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})] \\
&-\frac{f_1(\mathbf{x})}{\ell(\mathbf{x})f(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})] + \frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^2\} \\
&-\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^2\} \\
&-\frac{f_1^2(\mathbf{x}) - f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]\} \\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^3\}\Bigg)^2 \\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^3\}.
\end{aligned}
$$

By straightforward calculations, it can be obtained that

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] \;=\; & \frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^2\} \\
&+\frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^2\} \\
&-\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]\} \\
&-\frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]\}^2 \\
&-\frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\{\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]\}^2 \\
&+\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})] \\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^3\}.
\end{aligned}
$$

So, taking into account the results in Lemma 3.3, the variance of $\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)$ is:

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};0)] \;=\; & \frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0)] + \frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0)] \\
&-\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f^2(\mathbf{x})}\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0), \hat{g}_{2,\mathbf{H}}(\mathbf{x};0)] \\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x};0) - g_1(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x};0) - g_2(\mathbf{x})]^3\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)\nu_2^2(\mathbf{x}) + \frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)\nu_1^2(\mathbf{x}) \\
&\quad -\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)\tilde{c}(\mathbf{x}) + o\left(\frac{1}{n|\mathbf{H}|}\right) \\
&= \frac{1}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}R(K)[f_1^2(\mathbf{x})\nu_2^2(\mathbf{x}) + f_2^2(\mathbf{x})\nu_1^2(\mathbf{x}) - 2f_1(\mathbf{x})f_2(\mathbf{x})\tilde{c}(\mathbf{x})] \\
&\quad + o\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}
$$

Taking into account that $\nu_j^2(\mathbf{x}) = m_j(\mathbf{x}) + s_j^2(\mathbf{x}), j = 1, 2$, $\tilde{c}(\mathbf{x}) = m_1(\mathbf{x})m_2(\mathbf{x}) + c(\mathbf{x})$ and $f_1^2(\mathbf{x}) + f_2^2(\mathbf{x}) = 1$, and considering (3.8), (3.11), (3.12) and (3.13), it follows that

$$
\begin{aligned}
&f_1^2(\mathbf{x})\nu_2^2(\mathbf{x}) + f_2^2(\mathbf{x})\nu_1^2(\mathbf{x}) - 2f_1(\mathbf{x})f_2(\mathbf{x})\tilde{c}(\mathbf{x}) \\
&\quad = f_1^2(\mathbf{x})f_2^2(\mathbf{x})\ell^2(\mathbf{x}) + f_1^2(\mathbf{x})f_2^2(\mathbf{x})\sigma_2^2(\mathbf{x}) - 2f_2(\mathbf{x})f_1^3(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_1^4(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&\qquad + f_2^2(\mathbf{x})f_1^2(\mathbf{x})\ell^2(\mathbf{x}) + f_2^2(\mathbf{x})f_1^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_2^3(\mathbf{x})f_1(\mathbf{x})\sigma_{12}(\mathbf{x}) + f_2^4(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&\qquad - 2f_1^2(\mathbf{x})f_2^2(\mathbf{x})\ell^2(\mathbf{x}) - 2f_1^2(\mathbf{x})f_2^2(\mathbf{x})\sigma_2^2(\mathbf{x}) + 2f_1^3(\mathbf{x})f_2(\mathbf{x})\sigma_{12}(\mathbf{x}) \\
&\qquad - 2f_2^3(\mathbf{x})f_1(\mathbf{x})\sigma_{12}(\mathbf{x}) + 2f_1^2(\mathbf{x})f_2^2(\mathbf{x})\sigma_1^2(\mathbf{x}) \\
&\quad = \sigma_1^2(\mathbf{x}).
\end{aligned} \tag{3.61}
$$

Therefore,

$$
\mathbb{V}\mathrm{ar}[\tilde{m}_{\mathbf{H}}^c(\mathbf{x}; 0)] = \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o\left(\frac{1}{n|\mathbf{H}|}\right).
$$

Proof of Lemma 3.4 First, using the result for the conditional mean $\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x}; 1)] = \mathbb{E}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x}; 1)|\mathbf{X}_1, \ldots, \mathbf{X}_n]\}$, and that

$$
\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x}; 1)|\mathbf{X}_1, \ldots, \mathbf{X}_n] = \frac{1}{n}\sum_{i=1}^{n} m_j(\mathbf{X}_i)\hat{W}_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}; 1), \tag{3.62}
$$

with the weights $\hat{W}_{\mathbf{H}}(\mathbf{u}; 1)$ given in (3.31), under the assumptions (A1), (C1), (H1)

and (K1), and using (3.50), (3.51) and (3.52), it follows that,

$$
\begin{aligned}
&\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)] \\
&= \int m_j(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u} \\
&= \int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x})m_j(\mathbf{u})f(\mathbf{u})\{\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x}) + o[\mathrm{tr}(\mathbf{H}^2)] \\
&\quad -(\mathbf{u}-\mathbf{x})^{\mathrm{T}}[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x}) + o(\mathbf{H}^2\mathbf{1}_d)]\}d\mathbf{u} \\
&= \int K(\mathbf{p})m_j(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})\{\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})d\mathbf{p} + o[\mathrm{tr}(\mathbf{H}^2)]\} \\
&\quad -\int K(\mathbf{p})m_j(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})(\mathbf{H}\mathbf{p})^{\mathrm{T}}[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x}) + o(\mathbf{H}^2\mathbf{1}_d)]d\mathbf{p} \\
&= \int K(\mathbf{p})[m_j(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}] \\
&\quad \cdot[f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}]\{\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x}) + o[\mathrm{tr}(\mathbf{H}^2)]\}d\mathbf{p} \\
&\quad -\int K(\mathbf{p})[m_j(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}] \\
&\quad \cdot[f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x}) + \frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}]\mathbf{p}^{\mathrm{T}}\mathbf{H}[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x}) + o(\mathbf{H}^2\mathbf{1}_d)]d\mathbf{p} \\
&= \mu_2(K)\mathrm{tr}(\mathbf{H}^2)f^2(\mathbf{x})m_j(\mathbf{x})\int K(\mathbf{p})d\mathbf{p} \\
&\quad +\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x})d\mathbf{p} \\
&\quad +\frac{1}{2}\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})m_j(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} \\
&\quad +\frac{1}{2}\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f^2(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})\mathbf{H}\mathbf{p}d\mathbf{p} \\
&\quad -\mu_2(K)m_j(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}^3\boldsymbol{\nabla}f(\mathbf{x})d\mathbf{p} \\
&\quad -\mu_2(K)f(\mathbf{x})\int K(\mathbf{p})\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}m_j(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}^3\boldsymbol{\nabla}f(\mathbf{x})d\mathbf{p} + o[\mathrm{tr}(\mathbf{H}^4)] \\
&= \mu_2(K)\mathrm{tr}(\mathbf{H}^2)f^2(\mathbf{x})m_j(\mathbf{x}) + \mu_2^2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\nabla}m_j(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})] \\
&\quad +\frac{1}{2}\mu_2^2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})m_j(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_f(\mathbf{x})] + \frac{1}{2}\mu_2^2(K)\mathrm{tr}(\mathbf{H}^2)f^2(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m_j}(\mathbf{x})] \\
&\quad -\mu_2^2(K)m_j(\mathbf{x})\mathrm{tr}[\mathbf{H}^4\boldsymbol{\nabla}f(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})] - \mu_2^2(K)f(\mathbf{x})\mathrm{tr}[\mathbf{H}^4\boldsymbol{\nabla}f(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m_j(\mathbf{x})] + o[\mathrm{tr}(\mathbf{H}^4)] \\
&= \mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})g_j(\mathbf{x}) + \frac{1}{2}\mu_2^2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_j}(\mathbf{x})] \\
&\quad -\mu_2^2(K)\boldsymbol{\nabla}^{\mathrm{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_j(\mathbf{x}) + o[\mathrm{tr}(\mathbf{H}^4)].
\end{aligned}
$$

Concerning the variance of $\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)$, for $j=1,2$, it follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)] &= \mathbb{E}\{\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\ &\quad +\mathbb{V}\mathrm{ar}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}.\end{aligned}$$

Firstly,

$$\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]=\frac{1}{n^2}\sum_{i=1}^{n}s_j^2(\mathbf{X}_i)\hat{W}_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x};1). \tag{3.63}$$

Therefore, considering (3.31) and under the assumptions (A1), (C1), (H1) and (K1),

$$\begin{aligned}&\mathbb{E}\{\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\ &= \frac{1}{n}\int s_j^2(\mathbf{u})\hat{W}_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\\ &= \frac{1}{n}\int K_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x})s_j^2(\mathbf{u})f(\mathbf{u})\\ &\quad \cdot\left\{\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})+\boldsymbol{o}[\mathrm{tr}(\mathbf{H}^2)]-(\mathbf{u}-\mathbf{x})^{\mathrm{T}}[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x})+o(\mathbf{H}^2\mathbf{1}_d)]\right\}^2 d\mathbf{u}\\ &= \frac{1}{n|\mathbf{H}|}\int K^2(\mathbf{p})s_j^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})\{\mu_2^2(K)\mathrm{tr}^2(\mathbf{H}^2)f^2(\mathbf{x})+\boldsymbol{o}[\mathrm{tr}^2(\mathbf{H}^2)]\}d\mathbf{p}\\ &\quad +\frac{1}{n|\mathbf{H}|}\int K^2(\mathbf{p})s_j^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})\{\mathbf{p}^{\mathrm{T}}\mathbf{H}[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x})+o(\mathbf{H}^2\mathbf{1}_d)]\}^2 d\mathbf{p}\\ &\quad -\frac{2}{n|\mathbf{H}|}\int K^2(\mathbf{p})s_j^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})\mu_2(K)\mathrm{tr}(\mathbf{H}^2)f(\mathbf{x})\mathbf{p}^{\mathrm{T}}\mathbf{H}\\ &\quad \cdot[\mu_2(K)\mathbf{H}^2\boldsymbol{\nabla}f(\mathbf{x})+o(\mathbf{H}^2\mathbf{1}_d)]d\mathbf{p}\\ &= \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)s_j^2(\mathbf{x})f^3(\mathbf{x})\mathrm{tr}^2(\mathbf{H}^2)+\boldsymbol{o}\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right].\end{aligned} \tag{3.64}$$

On the other hand, using the expressions (3.31) and (3.62), it follows that under

the assumptions (A1), (C1), (H1) and (K1),

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\{\mathbb{E}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\
&= \frac{1}{n}\int m_j^2(\mathbf{u})\hat{W}_{\mathbf{H}}^2(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\\
&\quad -\frac{1}{n}\int m_j(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\int m_j(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};1)f(\mathbf{v})d\mathbf{v}\\
&= \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)m_j^2(\mathbf{x})f^3(\mathbf{x})\mathrm{tr}^2(\mathbf{H}^2)-\frac{1}{n}\mu_2^2(K)\mathrm{tr}^2(\mathbf{H}^2)f^2(\mathbf{x})g_j^2(\mathbf{x})\\
&\quad +o\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right]. \qquad (3.65)
\end{aligned}$$

Hence, from (3.63), (3.64), (3.65), and observing that $\nu_j^2(\mathbf{x}) = m_j^2(\mathbf{x}) + s_j^2(\mathbf{x})$, one gets that

$$\mathbb{V}\mathrm{ar}[\hat{g}_{j,\mathbf{H}}(\mathbf{x};1)] = \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)\nu_j^2(\mathbf{x})f^3(\mathbf{x})\mathrm{tr}^2(\mathbf{H}^2)+o\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right].$$

Finally, for the covariance, taking into account that

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};1),\hat{g}_{2,\mathbf{H}}(\mathbf{x};1)]\\
&= \mathbb{E}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};1)\hat{g}_{2,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\\
&\quad -\mathbb{E}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\}\mathbb{E}\{\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\},
\end{aligned}$$

and that

$$\begin{aligned}
&\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};1)\hat{g}_{2,\mathbf{H}}(\mathbf{x};1)|\mathbf{X}_1,\ldots,\mathbf{X}_n]\\
&= \frac{1}{n^2}\sum_{i=1}^{n}\tilde{c}(\mathbf{X}_i)\hat{W}_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x};1)\\
&\quad +\frac{1}{n^2}\sum_{i\neq j}^{n}m_1(\mathbf{X}_i)m_2(\mathbf{X}_j)\hat{W}_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x};1)\hat{W}_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x};1),
\end{aligned}$$

under assumptions (C1), using (3.31), and following similar arguments to those used for the bias and the variance, it follows that

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x};1),\hat{g}_{2,\mathbf{H}}(\mathbf{x};1)]\\
&= \frac{1}{n}\int \tilde{c}(\mathbf{u})\hat{W}^2_{\mathbf{H}}(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\\
&\quad +\frac{n-1}{n}\int m_1(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\int m_2(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};1)f(\mathbf{v})d\mathbf{v}\\
&\quad -\int m_1(\mathbf{u})\hat{W}_{\mathbf{H}}(\mathbf{u}-\mathbf{x};1)f(\mathbf{u})d\mathbf{u}\int m_2(\mathbf{v})\hat{W}_{\mathbf{H}}(\mathbf{v}-\mathbf{x};1)f(\mathbf{v})d\mathbf{v}\\
&= \frac{1}{n|\mathbf{H}|}\mu_2^2(K)R(K)\tilde{c}(\mathbf{x})f^3(\mathbf{x})\mathrm{tr}^2(\mathbf{H}^2)+o\left[\frac{\mathrm{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right].
\end{aligned}
$$

Proof of Theorem 3.6 For the bias, the function $\mathrm{atan2}(\hat{g}_{1,\mathbf{H}},\hat{g}_{2,\mathbf{H}})$ is expanded in Taylor series around $(g_1^*,g_2^*)=(\mathrm{tr}(\mathbf{H}^2)f\mu_2(K)g_1,\mathrm{tr}(\mathbf{H}^2)f\mu_2(K)g_2)$:

$$
\begin{aligned}
\mathrm{atan2}(\hat{g}_{1,\mathbf{H}},\hat{g}_{2,\mathbf{H}}) &= \mathrm{atan2}(g^*_{1,\mathbf{H}},g^*_{2,\mathbf{H}})+\frac{f_2}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell f^2}(\hat{g}_{1,\mathbf{H}}-g_1^*)\\
&\quad -\frac{f_1}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell f^2}(\hat{g}_{2,\mathbf{H}}-g_2^*)+\frac{f_1f_2}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2f^4}(\hat{g}_{2,\mathbf{H}}-g_2^*)^2\\
&\quad -\frac{f_1f_2}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2f^4}(\hat{g}_{1,\mathbf{H}}-g_1^*)^2\\
&\quad -\frac{f_1^2-f_2^2}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2f^4}(\hat{g}_{1,\mathbf{H}}-g_1^*)(\hat{g}_{2,\mathbf{H}}-g_2^*)\\
&\quad +\mathcal{O}[(\hat{g}_{1,\mathbf{H}}-g_1^*)^3]+\mathcal{O}[(\hat{g}_{2,\mathbf{H}}-g_2^*)^3]. \qquad (3.66)
\end{aligned}
$$

Hence, taking expectations in the above expression, it follows that

$$
\begin{aligned}
&\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};1)]-m^c(\mathbf{x})\\
&= \frac{f_2(\mathbf{x})}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]\\
&\quad -\frac{f_1(\mathbf{x})}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\\
&\quad +\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^2\}\\
&\quad -\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^2\}\\
&\quad -\frac{f_1^2(\mathbf{x})-f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\}\\
&\quad +\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^3\}+\mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^3\}.
\end{aligned}
$$

Noting that $\mathbb{E}[(\hat{g}_i - g_i)^2] = \mathbb{V}\text{ar}(\hat{g}_i) + [\mathbb{E}(\hat{g}_i) - g_i]^2$, and using the results in Lemma 3.4, one gets that

$$\begin{aligned}
&\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};1)] - m^c(\mathbf{x}) \\
&= \frac{f_2(\mathbf{x})}{\text{tr}(\mathbf{H}^2)\ell(\mathbf{x})f^2(\mathbf{x})}\mu_2(K)\left\{\frac{1}{2}\text{tr}(\mathbf{H}^2)f(\mathbf{x})\text{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_1}(\mathbf{x})] - \boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_1(\mathbf{x})\right\} \\
&\quad -\frac{f_1(\mathbf{x})}{\text{tr}(\mathbf{H}^2)\ell(\mathbf{x})f^2(\mathbf{x})}\mu_2(K)\left\{\frac{1}{2}\text{tr}(\mathbf{H}^2)f(\mathbf{x})\text{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{g_2}(\mathbf{x})] - \boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_2(\mathbf{x})\right\} \\
&\quad +O\left[\frac{\text{tr}^2(\mathbf{H}^2)}{n|\mathbf{H}|}\right] + o[\text{tr}(\mathbf{H}^2)].
\end{aligned}$$

Using (3.57) and (3.58), it can be obtained that

$$\begin{aligned}
&f_2(\mathbf{x})\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_1(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}g_2(\mathbf{x}) \\
&\quad = \boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4[f_2(\mathbf{x})\boldsymbol{\nabla}(\ell f)(\mathbf{x})f_1(\mathbf{x}) + f_2(\mathbf{x})\boldsymbol{\nabla}f_1(\mathbf{x})(\ell f)(\mathbf{x}) \\
&\qquad -f_1(\mathbf{x})\boldsymbol{\nabla}(\ell f)(\mathbf{x})f_2(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla}f_2(\mathbf{x})(\ell f)(\mathbf{x})] \\
&\quad = \boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4(\ell f)(\mathbf{x})[f_2(\mathbf{x})\boldsymbol{\nabla}f_1(\mathbf{x}) - f_1(\mathbf{x})\boldsymbol{\nabla}f_2(\mathbf{x})] \\
&\quad = (\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}m^c(\mathbf{x}).
\end{aligned}$$

Then, using (3.59) and the last expression, it follows that

$$\begin{aligned}
&\mathbb{E}[\tilde{m}^c_{\mathbf{H}}(\mathbf{x};1)] - m^c(\mathbf{x}) \\
&= \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})}\frac{1}{2}\{2\boldsymbol{\nabla}^{\text{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x}) + \text{tr}[\ell(\mathbf{x})f(\mathbf{x})\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})]\} \\
&\quad -\frac{\mu_2(K)}{\text{tr}(\mathbf{H}^2)f(\mathbf{x})}\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}m^c(\mathbf{x}) + o[\text{tr}(\mathbf{H}^2)] \\
&= \frac{\mu_2(K)}{\ell(\mathbf{x})}\boldsymbol{\nabla}^{\text{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}\ell(\mathbf{x}) + \frac{\mu_2(K)}{f(\mathbf{x})}\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}m^c(\mathbf{x}) \\
&\quad +\frac{\mu_2(K)}{2}\text{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] - \frac{\mu_2(K)}{\text{tr}(\mathbf{H}^2)f(\mathbf{x})}\boldsymbol{\nabla}^{\text{T}}f(\mathbf{x})\mathbf{H}^4\boldsymbol{\nabla}m^c(\mathbf{x}) + o[\text{tr}(\mathbf{H}^2)] \\
&= \frac{1}{2}\mu_2(K)\text{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})}\boldsymbol{\nabla}^{\text{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}\ell(\mathbf{x}) + o[\text{tr}(\mathbf{H}^2)].
\end{aligned}$$

To derive the variance, the function $\text{atan2}^2(\hat{g}_{1,\mathbf{H}}, \hat{g}_{2,\mathbf{H}})$ is expanded in Taylor series around $(g_1^*, g_2^*) = (\text{tr}(\mathbf{H}^2) f \mu_2(K) g_1, \text{tr}(\mathbf{H}^2) f \mu_2(K) g_2)$.

$$\begin{aligned}
\text{atan2}^2(\hat{g}_{1,\mathbf{H}}, \hat{g}_{2,\mathbf{H}}) &= \text{atan2}^2(g_{1,\mathbf{H}}^*, g_{2,\mathbf{H}}^*) + \frac{2\text{atan2}(g_1, g_2) f_2}{\text{tr}(\mathbf{H}^2)\mu_2(K)\ell f^2}(\hat{g}_{1,\mathbf{H}} - g_1^*) \\
&\quad - \frac{2\text{atan2}(g_1, g_2) f_1}{\text{tr}(\mathbf{H}^2)\mu_2(K)\ell f^2}(\hat{g}_{2,\mathbf{H}} - g_2^*) \\
&\quad + \frac{2\text{atan2}(g_1, g_2) f_1 f_2}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{2,\mathbf{H}} - g_2^*)^2 \\
&\quad - \frac{2\text{atan2}(g_1, g_2) f_1 f_2}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{1,\mathbf{H}} - g_1^*)^2 \\
&\quad - \frac{2\text{atan}(g_1, g_2)(f_1^2 - f_2^2)}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{1,\mathbf{H}} - g_1^*)(\hat{g}_{2,\mathbf{H}} - g_2^*) \\
&\quad + \frac{f_1^2}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{2,\mathbf{H}} - g_2^*)^2 \\
&\quad + \frac{f_2^2}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{1,\mathbf{H}} - g_1^*)^2 \\
&\quad - \frac{2 f_1 f_2}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2 f^4}(\hat{g}_{1,\mathbf{H}} - g_1^*)(\hat{g}_{2,\mathbf{H}} - g_2^*) \\
&\quad + \mathcal{O}[(\hat{g}_{1,\mathbf{H}} - g_1^*)^3] + \mathcal{O}[(\hat{g}_{2,\mathbf{H}} - g_2^*)^3].
\end{aligned}$$

Now, taking expectations in the above expression and in the expression (3.66), it is obtained that

$$\begin{aligned}
&\mathbb{V}\text{ar}[\tilde{m}_{\mathbf{H}}^c(\mathbf{x}; 1)] \\
&= \text{atan2}^2[g_{1,\mathbf{H}}(\mathbf{x}), g_{2,\mathbf{H}}(\mathbf{x})] + \frac{2\text{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_2(\mathbf{x})}{\text{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x}) - g_1^*(\mathbf{x})] \\
&\quad - \frac{2\text{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x})}{\text{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x}) f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x}) - g_2^*(\mathbf{x})] \\
&\quad + \frac{2\text{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x}) f_2(\mathbf{x})}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x}) f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x}) - g_2^*(\mathbf{x})]^2\} \\
&\quad - \frac{2\text{atan2}[g_1(\mathbf{x}), g_2(\mathbf{x})] f_1(\mathbf{x}) f_2(\mathbf{x})}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x}) f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}) - g_1^*(\mathbf{x})]^2\} \\
&\quad - \frac{2\text{atan}[g_1(\mathbf{x}), g_2(\mathbf{x})][f_1^2(\mathbf{x}) - f_2^2(\mathbf{x})]}{\text{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x}) f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}) - g_1^*(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x}) - g_2^*(\mathbf{x})]\}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{f_1^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^2\}\\
&+\frac{f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^2\}\\
&-\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\}\\
&-\Bigg(\mathrm{atan2}[g_1(\mathbf{x}),g_2(\mathbf{x})]+\frac{f_2(\mathbf{x})}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]\\
&-\frac{f_1(\mathbf{x})}{\mathrm{tr}(\mathbf{H}^2)\mu_2(K)\ell(\mathbf{x})f^2(\mathbf{x})}\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\\
&+\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^2\}\\
&-\frac{f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^2\}\\
&-\frac{f_1^2(\mathbf{x})-f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\}\\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^3\}+\mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^3\}\Bigg)^2\\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^3\}+\mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^3\}.
\end{aligned}
$$

By straightforward calculations, it can be obtained that

$$
\begin{aligned}
\mathbb{V}\mathrm{ar}[\tilde{m}_{\mathbf{H}}^c(\mathbf{x};1)] \;=\;& \frac{f_1^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^2\}\\
&+\frac{f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^2\}\\
&-\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})][\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\}\\
&-\frac{f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\{\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]\}^2\\
&-\frac{f_1^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\{\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\}^2\\
&+\frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{E}[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]\mathbb{E}[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]\\
&+\mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x})-g_1^*(\mathbf{x})]^3\}+\mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x})-g_2^*(\mathbf{x})]^3\}.
\end{aligned}
$$

So, taking into account the results of Lemma 3.4 and considering (3.61), the variance of $\tilde{m}^{c}_{\mathbf{H}}(\mathbf{x};1)$ is:

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\tilde{m}^{c}_{\mathbf{H}}(\mathbf{x};1)] &= \frac{f_1^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{g}_{2,\mathbf{H}}(\mathbf{x})] \\
&\quad + \frac{f_2^2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{g}_{1,\mathbf{H}}(\mathbf{x})] \\
&\quad - \frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\mathrm{tr}^2(\mathbf{H}^2)\mu_2^2(K)\ell^2(\mathbf{x})f^4(\mathbf{x})}\mathbb{C}\mathrm{ov}[\hat{g}_{1,\mathbf{H}}(\mathbf{x}),\hat{g}_{2,\mathbf{H}}(\mathbf{x})] \\
&\quad + \mathcal{O}\{[\hat{g}_{1,\mathbf{H}}(\mathbf{x}) - g_1^*(\mathbf{x})]^3\} + \mathcal{O}\{[\hat{g}_{2,\mathbf{H}}(\mathbf{x}) - g_2^*(\mathbf{x})]^3\} \\
&= \frac{f_1^2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)\nu_2^2(\mathbf{x}) + \frac{f_2^2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)\nu_1^2(\mathbf{x}) \\
&\quad - \frac{2f_1(\mathbf{x})f_2(\mathbf{x})}{\ell^2(\mathbf{x})f(\mathbf{x})}\frac{1}{n|\mathbf{H}|}R(K)c(\mathbf{x}) + o\left(\frac{1}{n|\mathbf{H}|}\right) \\
&= \frac{1}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}R(K)[f_1^2(\mathbf{x})\nu_2^2(\mathbf{x}) + f_2^2(\mathbf{x})\nu_1^2(\mathbf{x}) - 2f_1(\mathbf{x})f_2(\mathbf{x})c(\mathbf{x})] \\
&\quad + o\left(\frac{1}{n|\mathbf{H}|}\right) \\
&= \frac{R(K)\sigma_1^2(\mathbf{x})}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

Chapter 4

Nonparametric regression estimation for a circular response and an $\mathbb{R}^d$-valued covariate with spatially correlated errors

4.1 Introduction

In many scientific fields, such as oceanography, meteorology or biology, data are angular measurements (points in the unit circle), exhibiting in some cases a spatial dependence structure which should be accounted for in any modeling approach. Literature on circular spatial modeling is not very extensive. For geostatistical data, Morphet (2009) introduced the cosineogram (trying to imitate the variogram) as a measure of spatial autocorrelation between angles. The author also presented a method to simulate spatially correlated circular data using a transformation of a Gaussian spatial process. For the analysis of wind data (speed and direction), Modlin et al. (2012) proposed a Bayesian hierarchical model, specified by a circular conditional autoregressive process. This model is based on the wrapped circular distribution of the direction vector and a spatial Gaussian autoregressive model for the logarithm of the wind speed. In a practical setting, the authors analyzed hurricane surface wind fields. In a similar framework, Casson and Coles (1998) provided a

spatial analysis about the direction of maximum wind speed at locations on the Gulf and Atlantic coasts of the United States, by considering conditionally independent directions modeled with a von Mises distribution, introducing the spatial structure in the modal direction and concentration parameters.

In other scenarios, circular measurements are also accompanied by observations of other real-valued random variables, as in Mastrantonio et al. (2018), who proposed a Markov model for multivariate circular-linear data to forecast the wind speed and direction in the city of Taranto (Italy); or García-Portugués et al. (2014), who analyzed the relation between orientation and size of wildfires in Portugal. Alternative approaches using copulas for dependence modeling have been also considered in similar contexts. For instance, Carnicero et al. (2013), explored the relation between wind direction and rainfall amount in the North of Spain, as well as the dependence between the wind directions in two nearby buoys at the Atlantic ocean.

Focusing on the analysis of wave heights and wave directions, some other authors developed different approaches to model spatial or spatio-temporal circular data. For example, Jona-Lasinio et al. (2012) formulated the so-called wrapped Gaussian spatial process, as a spatial process for circular data, allowing only symmetric marginal distributions. They analyzed outgoing wave directions at the Adriatic Sea area during a storm period by using a Markov Chain Monte Carlo (MCMC) model fitting. Mastrantonio et al. (2016) introduced the wrapped skew Gaussian spatial process as an alternative to the wrapped Gaussian spatial process which allows for asymmetric marginal distributions. This circular process was also used to analyze wave directions at the Adriatic Sea. Wang and Gelfand (2014) developed the projected Gaussian spatial process, induced from a linear bivariate Gaussian spatial process. Using MCMC methods, the authors modeled wave directions at Adriatic Sea from a calm period transitioning to a storm period. Wang et al. (2015) proposed a fully model-based approach to capture spatial and temporal joint dependence structure between a linear and an angular variable. More specifically, the joint distribution of the wave height and direction is specified by modeling the conditional distribution of the wave height given the wave direction through a Bayesian geostatistical model. Lagona et al. (2015) introduced a hidden Markov model accounting for the correlation of spatio-temporal linear-circular data, providing an approach to iden-

tify regimes of marine currents. The author considered a flexible model based on the multivariate von Mises distribution, adapting it to heteroscedasticity and an autoregressive correlation structure.

An alternative to the previous approaches to analyze spatial circular data consists in the formulation of a linear-circular regression model taking the spatial correlation into account and, subsequently, estimating the corresponding circular regression function. Following this idea and for a single real-valued covariate, Di Marzio et al. (2013) introduced a nonparametric estimator of the regression function when the data come from mixing processes. The same approach has been also applied, with due modifications, in the context of time series by Di Marzio et al. (2012). The authors considered smoothing and prediction in the time domain for circular time-series data. In this chapter, using a similar approach, the regression model (3.3) with circular response and several Euclidean covariates, but assuming spatially correlated errors, is considered. Then, the nonparametric estimators given in (3.14) are analyzed in this framework. As expected, the asymptotic variance of these estimators will depend on the correlation.

As pointed out in the previous chapter, a crucial step to compute estimators (3.14) is the selection of an appropriate bandwidth or smoothing matrix. Following similar arguments to those used in Chapter 3, some guidelines to select locally optimal bandwidth matrices for the nonparametric estimators given in (3.14) are provided in this chapter, when considering the regression model (3.3) for spatially correlated errors. For practical purposes, cross-validation criteria are also defined and analyzed in practice. It should be noted that for correlated data, the standard leave-one-out cross-validation method will not provide satisfactory bandwidths (see Section 1.1.2 for further details). In this chapter, a modified version of the cross-validation criterion (MCV^c), given in (3.33), is provided. Different simulation scenarios will be designed, considering circular spatial errors generated from wrapped and projected Gaussian processes, introduced in Jona-Lasinio et al. (2012) and Wang and Gelfand (2014), respectively.

This chapter is organized as follows. An introduction on different models of spatial processes for circular data are provided in Section 4.2. Some of these models will be employed in the simulation study included in Section 4.5. Section 4.3 intro-

duces the linear-circular regression model, which adds a spatial random process to a deterministic trend. Marginal models for the sine and cosine components required for the proposed estimators are also presented in this section. In Section 4.4, the asymptotic conditional variance of the nonparametric estimators of the circular regression function, given in (3.14), are derived, under certain conditions on the model dependence structure. Additionally, some proposals for bandwidth selection are introduced. A simulation study for assessing the performance of the estimators and the bandwidth selectors, in this context of spatial dependence, is provided in Section 4.5. The simulations were carried out considering that the errors are drawn from a wrapped and a projected Gaussian spatial process. Section 4.6 shows the illustration of the nonparametric approaches to the wave direction dataset (for a certain time and period) in the Adriatic Sea area previously cited. Finally, Section 4.7 includes the proofs of the main results.

4.2 Modeling circular processes with spatial dependence

This section describes different models of spatial processes for circular data. Some of these models will be employed in the simulation study presented in Section 4.5. In Section 4.2.1, we briefly describe the wrapped Gaussian spatial process (Jona-Lasinio et al., 2012), which is induced from a linear Gaussian spatial process. Projected Gaussian spatial processes (Wang and Gelfand, 2014), which are obtained from linear bivariate Gaussian spatial processes, are presented in Section 4.2.2. Other circular spatial processes such as those obtained from transformations of Gaussian spatial processes (Morphet, 2009), are formulated in Section 4.2.3.

4.2.1 Wrapped Gaussian spatial processes

Wrapped Gaussian spatial processes, which are induced from linear Gaussian spatial processes, were introduced by Jona-Lasinio et al. (2012). The wrapping approach consists on *wrapping* a linear variable around the unit circle, and consequently, its circular density function is obtained by using a modulo operation and wrapping the

density function of the linear random variable. Next, the steps followed to construct a wrapped Gaussian spatial process are described.

Given a collection of spatial coordinates, $\mathbf{X}_i$, with $i = 1, \ldots, n$, consider a realization $\{Y_i = Y(\mathbf{X}_i),\ i = 1, \ldots, n\}$ from a real-valued Gaussian spatial process, where each observation can be decomposed as:

$$Y_i = \mu + w_i, \quad i = 1, \ldots, n \tag{4.1}$$

being $\mu = \mu(\mathbf{X}_i)$ the mean and w_i random variables of a zero mean Gaussian spatial process with $\mathbb{C}\mathrm{ov}(w_i, w_j \mid \mathbf{X}_i, \mathbf{X}_j) = \sigma^2 \rho_n(\mathbf{X}_i - \mathbf{X}_j)$. The variance of w_i is denoted by σ^2 and ρ_n is a continuous stationary correlation function satisfying $\rho_n(\mathbf{0}) = 1$, $\rho_n(\mathbf{x}) = \rho_n(-\mathbf{x})$, and $|\rho_n(\mathbf{x})| \leq 1$, $\forall \mathbf{x}$. Then, a realization of a wrapped spatial process $\{\Theta_i,\ i = 1, \ldots, n\}$, linked to the spatial coordinates $\mathbf{X}_i$, with $i = 1, \ldots, n$, is obtained as:

$$\Theta_i = Y_i (\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n.$$

Note that this realization can be viewed in vector form as $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_n)^{\mathrm{T}}$, with mean direction vector $\mu \mathbf{1}_n$, being $\mathbf{1}_n$ a $n \times 1$ vector with every entry equal to 1, and covariance matrix $\sigma^2 \mathbf{R}_n$, where $\mathbf{R}_n(i, j) = \rho_n(\mathbf{X}_i - \mathbf{X}_j)$ is the (i, j)-entry of the correlation matrix $\mathbf{R}_n$.

The exponential spatial correlation function

$$\mathbb{C}\mathrm{ov}(w_i, w_j \mid \mathbf{X}_i, \mathbf{X}_j) = \sigma^2 [\exp(-\|\mathbf{X}_i - \mathbf{X}_j\| / a_e)], \tag{4.2}$$

where a_e is the practical range, and the circular correlation of the corresponding wrapped Gaussian spatial process (using a circular correlation coefficient) were compared in Jona-Lasinio et al. (2012, Figure 4), obtaining very similar shapes for both correlations. This comparison is also performed here using a simulated dataset. For this purpose, 500 samples of size $n = 400$ are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, from a Gaussian spatial process with zero mean and exponential variogram (4.2) with $\sigma^2 = 1$ and two values of the range parameter a_e ($a_e = 0.1, 0.3$). The corresponding zero mean wrapped spatial processes with the same values of σ^2 and a_e are computed. Figure 4.1 shows the exponential correla-

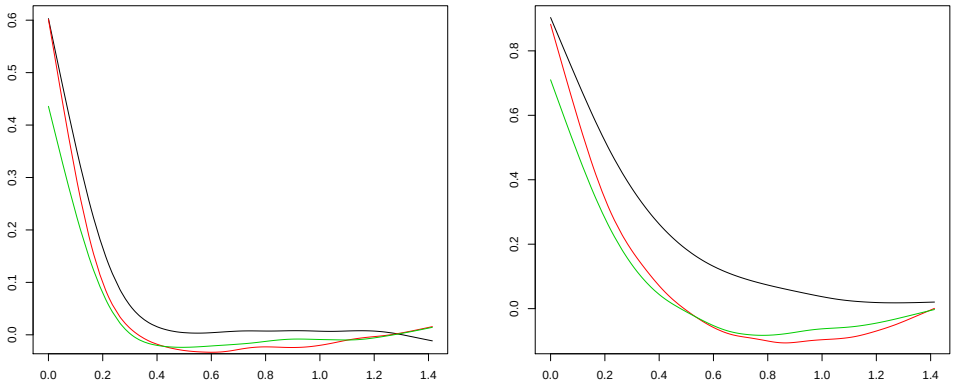

Figure 4.1: Exponential correlation function (black lines) and the corresponding correlation coefficient r_n at different distances, for a wrapped (red lines) and a projected Gaussian spatial process (green lines), for two values of the range parameter, $a_e = 0.1$ (left panel) and $a_e = 0.3$ (right panel) for $\mu = 0$ and $\sigma = 1$.

tion function (black lines) and the associated circular sample correlation coefficient r_n (red lines) given in (1.39), for two values of the range parameter, $a_e = 0.1$ (left panel) and $a_e = 0.3$ (right panel). As a_e gets larger, the circular correlation can depart considerably from the exponential correlation function of the inline process. However, the simulation results show that the circular correlation function of the wrapped Gaussian spatial process is monotonically decreasing as the distance between two locations increases, and the induced circular correlation curve is always below the spatial correlation curve for the linear process, except for large distances. More accurate results could possibly be obtained if the sample was larger.

Fixing the values of $\mu = 0$ and $\sigma = 1$, the effect of the range parameter a_e on a realization on a 15×15 grid of the wrapped circular process can be seen in Figure 4.2. Larger values of the range a_e yield a smoother pattern. Similar arguments were pointed out in Jona-Lasinio et al. (2012).

4.2.2 Projected Gaussian spatial processes

Another way of generating circular observations that exhibit spatial correlation are the projected Gaussian spatial processes, which are induced from bivariate Gaussian spatial processes (Wang and Gelfand, 2014).

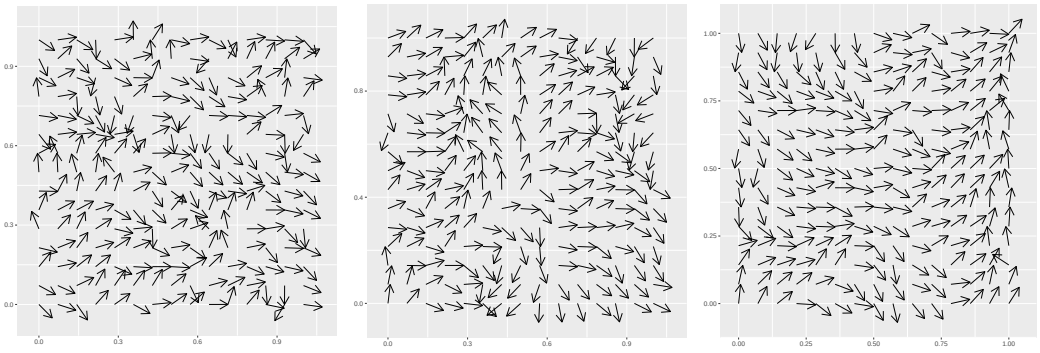

Figure 4.2: Simulated samples of a wrapped Gaussian spatial process on a 15×15 grid with exponential correlation being $a_e = 0.1$ (left), $a_e = 0.3$ (center) and $a_e = 0.6$ (right), for $\mu = 0$ and $\sigma = 1$ in (4.1) and (4.2).

Consider a bivariate Gaussian spatial process, $\mathbf{Y}$, observed at a collection of spatial coordinates $\mathbf{X}_i$, with $i = 1, \ldots, n$. The observations $\mathbf{Y}_i = (Y_{1i}, Y_{2i})$, with $\mathbf{Y}_i = \mathbf{Y}(\mathbf{X}_i)$ and $(Y_{1i}, Y_{2i}) = [Y_{1i}(\mathbf{X}_i), Y_{2i}(\mathbf{X}_i)]$, can be decomposed as follows:

$$\mathbf{Y}_i = \boldsymbol{\mu} + \mathbf{w}_i, \quad i = 1, \ldots, n, \tag{4.3}$$

where $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{X}_i) \in \mathbb{R}^2$ is the mean vector and the $\mathbf{w}_i$ are random variables of a zero mean bivariate Gaussian spatial process with cross covariance function $\rho_n(\mathbf{x}) \otimes \mathbf{T}$, being ρ_n a continuous stationary correlation function and $\mathbf{T}$ a matrix , defined as

$$\mathbf{T} = \begin{pmatrix} \sigma^2 & \tau\sigma \\ \tau\sigma & 1 \end{pmatrix},$$

with $\sigma > 0$ and $\tau \in [-1, 1]$. The operator $\otimes$ denotes the usual Kronecker product.

A realization of a circular spatial process (in vector form), $\boldsymbol{\Theta} = (\Theta_1, \ldots, \Theta_n)^{\mathrm{T}}$ can be obtained as:

$$\Theta_i = \operatorname{atan2}(Y_{2i}, Y_{1i}), \quad i = 1, \ldots, n.$$

Wang and Gelfand (2014) proved that the correlation functions of the projected and the original bivariate Gaussian spatial processes are *similar*. An analogous experiment to that perform in the previous section allows to compare the shape of

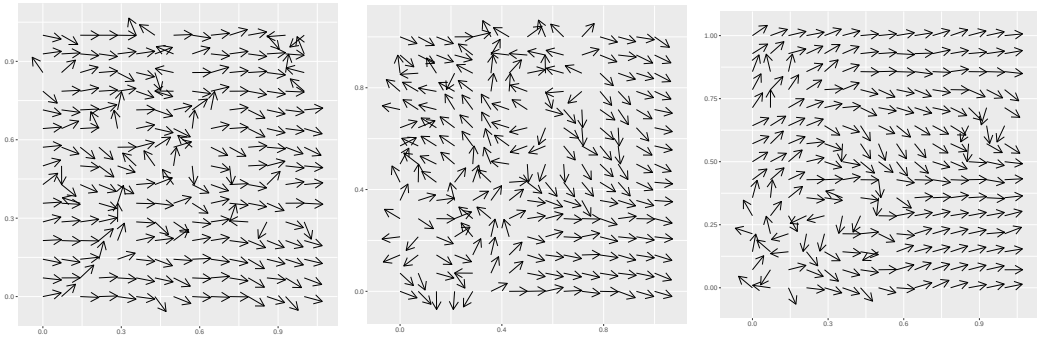

Figure 4.3: Simulated samples of a projected Gaussian spatial process on a 15×15 grid with exponential correlation being $a_e = 0.1$ (left), $a_e = 0.3$ (center) and $a_e = 0.6$ (right), for $\boldsymbol{\mu} = (1, 1)^{\mathrm{T}}$, $\sigma = 1$ and $\tau = 0.9$ in (4.3) and (4.4).

the correlation functions of the projected and the original bivariate Gaussian spatial processes. Considering $\boldsymbol{\mu} = (1, 1)^{\mathrm{T}}$ and the cross-covariance function

$$\mathbb{C}\mathrm{ov}(\mathbf{w}_i, \mathbf{w}_j \mid \mathbf{X}_i, \mathbf{X}_j) = [\exp(-\|\mathbf{X}_i - \mathbf{X}_j\|/a_e)] \cdot \mathbf{T}, \tag{4.4}$$

with $\sigma = 1$ and $\tau = 0.9$, Figure 4.1 shows the exponential correlation function (black line) and the corresponding correlation for the projected Gaussian spatial process (green line), computed at different distances, for two values of the range parameter, $a_e = 0.1$ (left panel) and $a_e = 0.3$ (right panel). Sensitivity to the choice of the larger range parameter a_e can be observed.

Figure 4.3 shows a sample on a 15×15 grid of a simulated projected Gaussian spatial process for different values of the range parameter, $a_e = 0.1$ (weak correlation), $a_e = 0.3$ (medium correlation) and $a_e = 0.6$ (strong correlation), for $\boldsymbol{\mu} = (1, 1)^{\mathrm{T}}$, $\sigma = 1$ and $\tau = 0.9$. From left to right, the range increases, i.e., there is a stronger spatial dependence structure, and, consequently, the corresponding circular process realization shows a *smoother* pattern.

4.2.3 Other circular spatial processes

In this section, other remarkable circular processes with spatial dependence are presented. More specifically, circular spatial processes obtained by transformations of Gaussian spatial processes (Morphet, 2009) are introduced. They represent an al-

ternative to generate circular data with spatial correlation, however, for the sake of illustration, wrapped and projected spatial processes (which were described in Sections 4.2.1 and 4.2.2, respectively) will be only employed in the simulation study of Section 4.5. Next, the steps followed to generate a sample of circular spatial processes using this approach are described.

Given a collection of spatial coordinates, $\mathbf{X}_i$, with $i = 1, \ldots, n$, generate a Gaussian spatial process $\{Y_i = Y(\mathbf{X}_i),\ i = 1, \ldots, n\}$ with mean $\mu = \mu(\mathbf{X}_i)$ and $\mathbb{C}\text{ov}(Y_i, Y_j \mid \mathbf{X}_i, \mathbf{X}_j) = \sigma^2 \rho_n(\mathbf{X}_i - \mathbf{X}_j)$. The variance of $Y_i, i = 1, \ldots, n$, is denoted by σ^2 and ρ_n is a continuous stationary correlation function satisfying $\rho_n(\mathbf{0}) = 1$, $\rho_n(\mathbf{x}) = \rho_n(-\mathbf{x})$, and $|\rho_n(\mathbf{x})| \leq 1$, $\forall \mathbf{x}$. Then, a realization of the circular spatial process $\{\Theta_i,\ i = 1, \ldots, n\}$, can be obtained by computing the inverse of the cumulative distribution function as:

$$\Theta_i = G_\Theta^{-1}[F_Y(Y_i)], \quad i = 1, \ldots, n,$$

where G_Θ and F_Y are the cumulative distribution functions of Θ and Y, respectively. Close expressions of the cumulative distribution functions for circular distributions, G_Θ, and their corresponding inverses were provided in Morphet (2009, Table 5.2). An example of a simulated circular spatial process with the von Mises distribution, obtained from a transformation of a Gaussian spatial process, can be found in Morphet (2009, Figure 5.1). We refer to Morphet (2009, Section 5.3) for further details on this method of generating a circular spatial process. Moreover, the distributional and spatial properties of this circular spatial process were derived by Morphet (2009, Section 5.4). For instance, the similarity on the shape of the inverted cosineogram (which is an analogue of the covariogram) of the circular spatial process and the variogram of a Gaussian spatial process was illustrated (Morphet, 2009, Figure 5.4).

4.3 Nonparametric circular regression estimation with spatially correlated errors

This section introduces the linear-circular regression model, which adds a spatial circular random process to the trend or regression function. In this case, we assume

the linear-circular regression model given in (3.3), but supposing that the circular errors are spatially correlated. More specifically, we consider the regression model given in (3.3):

$$\Theta_i = [m^c(\mathbf{X}_i) + \varepsilon_i](\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n, \tag{4.5}$$

where m^c is a smooth trend or regression function and the ε_i are random angles, such that, $\mathbb{E}[\sin(\varepsilon_i) \mid \mathbf{X} = \mathbf{x}] = 0$, and additionally, satisfying in this dependence framework that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}[\sin(\varepsilon_i), \sin(\varepsilon_j) \mid \mathbf{X}_i, \mathbf{X}_j] &= \sigma_1^2 \rho_{1,n}(\mathbf{X}_i - \mathbf{X}_j), \\
\mathbb{C}\mathrm{ov}[\cos(\varepsilon_i), \cos(\varepsilon_j) \mid \mathbf{X}_i, \mathbf{X}_j] &= \sigma_2^2 \rho_{2,n}(\mathbf{X}_i - \mathbf{X}_j), \\
\mathbb{C}\mathrm{ov}[\sin(\varepsilon_i), \cos(\varepsilon_j) \mid \mathbf{X}_i, \mathbf{X}_j] &= \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j),
\end{aligned}$$

with $\sigma_k^2 < \infty$ for $k = 1, 2$, and $\sigma_{12} < \infty$. The continuous stationary correlation functions $\rho_{k,n}$ satisfy $\rho_{k,n}(\mathbf{0}) = 1$, $\rho_{k,n}(\mathbf{x}) = \rho_{k,n}(-\mathbf{x})$, and $|\rho_{k,n}(\mathbf{x})| \leq 1$, for any $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$, and $k = 1, 2, 3$. The subscript n in $\rho_{k,n}$ indicates that the correlation functions vary with n (specifically, the correlation functions shrink as n goes to infinity). Note also that the subscript k does not correspond to an integer sequence and it just indicates if the correlation corresponds to the sine process ($k = 1$), the cosine process ($k = 2$) or if it is the cross-correlation between them ($k = 3$).

In this context, considering model (4.5), the regression function m^c can be nonparametrically estimated using the local polynomial type estimators given in (3.14). As pointed out in Chapter 3, this approach implicitly brings the consideration of two regression models for the sine and cosine components of Θ on $\mathbf{X}$. However, in contrast to models (3.4) and (3.5), in this case, the errors of these two models are also spatially correlated. These regression models are:

$$\sin(\Theta_i) = m_1(\mathbf{X}_i) + \xi_i \tag{4.6}$$

$$\cos(\Theta_i) = m_2(\mathbf{X}_i) + \zeta_i, \tag{4.7}$$

where, as in models (3.4) and (3.5), the ξ_i and the ζ_i, $i = 1, \ldots, n$, are samples of the variables ξ and ζ, respectively, absolutely bounded by 1, satisfying $\mathbb{E}(\xi \mid \mathbf{X} = \mathbf{x}) = \mathbb{E}(\zeta \mid \mathbf{X} = \mathbf{x}) = 0$. Additionally, for every $\mathbf{x} \in \mathcal{D}$, set $\mathbb{V}\mathrm{ar}(\xi \mid$

$\mathbf{X} = \mathbf{x}) = s_1^2(\mathbf{x})$, $\mathbb{V}\mathrm{ar}(\zeta \mid \mathbf{X} = \mathbf{x}) = s_2^2(\mathbf{x})$, $\mathbb{E}(\xi\zeta \mid \mathbf{X} = \mathbf{x}) = c(\mathbf{x})$, and taking into account that the errors in model (4.5) are spatially correlated, we use the notation $\mathbb{C}\mathrm{ov}(\xi_i, \xi_j \mid \mathbf{X}_i, \mathbf{X}_j) = C_{n,1}(\mathbf{X}_i, \mathbf{X}_j)$, $\mathbb{C}\mathrm{ov}(\zeta_i, \zeta_j \mid \mathbf{X}_i, \mathbf{X}_j) = C_{n,2}(\mathbf{X}_i, \mathbf{X}_j)$ and $\mathbb{C}\mathrm{ov}(\xi_i, \zeta_j \mid \mathbf{X}_i, \mathbf{X}_j) = C_{n,3}(\mathbf{X}_i, \mathbf{X}_j)$, for $i, j = 1, \ldots, n$, and $i \neq j$.

Notice that expressions (3.8), (3.11), (3.12) and (3.13) also hold in this framework. Moreover, recalling that $f_1(\mathbf{x}) = \sin[m^c(\mathbf{x})]$ and $f_2(\mathbf{x}) = \cos[m^c(\mathbf{x})]$, and using sine and cosine addition formulas, the following relation between the covariance $C_{n,1}$, $C_{n,2}$ and $C_{n,3}$, defined in models (4.6) and (4.7), and the correlations $\rho_{k,n}$, $k = 1, 2, 3$, directly derived from model (4.5), can be obtained:

$$\begin{aligned}
C_{n,1}(\mathbf{X}_i, \mathbf{X}_j) &= f_1(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_2^2 \rho_{2,n}(\mathbf{X}_i - \mathbf{X}_j) + f_1(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad + f_2(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad + f_2(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_1^2 \rho_{1,n}(\mathbf{X}_i - \mathbf{X}_j), \qquad (4.8) \\
C_{n,2}(\mathbf{X}_i, \mathbf{X}_j) &= f_2(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_2^2 \rho_{2,n}(\mathbf{X}_i - \mathbf{X}_j) - f_2(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad - f_1(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad + f_1(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_1^2 \rho_{1,n}(\mathbf{X}_i - \mathbf{X}_j), \qquad (4.9) \\
C_{n,3}(\mathbf{X}_i, \mathbf{X}_j) &= f_1(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_2^2 \rho_{2,n}(\mathbf{X}_i - \mathbf{X}_j) - f_1(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad + f_2(\mathbf{X}_i) f_2(\mathbf{X}_j) \sigma_{12} \rho_{3,n}(\mathbf{X}_i - \mathbf{X}_j) \\
&\quad - f_2(\mathbf{X}_i) f_1(\mathbf{X}_j) \sigma_1^2 \rho_{1,n}(\mathbf{X}_i - \mathbf{X}_j). \qquad (4.10)
\end{aligned}$$

In Chapter 3 (Section 3.3.1) of this dissertation, the asymptotic conditional bias and variance of the circular regression estimators, given in (3.14), were derived for independent data. In the following section, these asymptotic expressions for estimators (3.14), with polynomial degrees $p = 0$ and $p = 1$, considering the regression model with spatially correlated errors (4.5), will be obtained.

Note that the asymptotic conditional bias of these estimators is the same for independent and for dependent data and, therefore, in this chapter, we will only focus on deriving the their asymptotic conditional variance under model (4.5).

4.4 Asymptotic properties of local polynomial estimators

The Nadaraya–Watson and local linear estimators of the regression functions m_j (denoted by $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$), $j=1,2$, at $\mathbf{x}\in\mathcal{D}\subset\mathbb{R}^d$, were defined in (3.15) and (3.22), for $p=0$ and $p=1$, respectively. Asymptotic properties of $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$, $j=1,2, p=0,1$, considering models (4.6) and (4.7) for the sine and cosine components, respectively, can be obtained using some results given in Liu (2001). For this, apart from the hypotheses required in the case of independent data (Chapter 3), the following extra assumption on the design is needed:

(C2) For the correlation functions $\rho_{k,n}$, $k=1,2,3$, there exist ρ_{M_k} and ρ_{c_k} such that $n\int|\rho_{k,n}(\mathbf{x})|d\mathbf{x}<\rho_{M_k}$ and $\lim_{n\to\infty} n\int\rho_{k,n}(\mathbf{x})d\mathbf{x}=\rho_{c_k}$. For any sequence $\epsilon_n>0$ satisfying $n^{1/2}\epsilon_n\to\infty$,

$$n\int_{\|\mathbf{x}\|\geq\epsilon_n}|\rho_{k,n}(\mathbf{x})|d\mathbf{x}\to 0 \quad \text{as} \quad n\to\infty.$$

Note that hypothesis (C2) establishes similar conditions on the correlation functions to those specified in assumption (A3), included in the Introduction of this dissertation. Therefore, the comments pointed out in Section 1.1.2 regarding assumption (A3) can be adapted here for assumption (C2). Instead of the correlation function ρ_n, they refer to the correlation functions $\rho_{k,n}$, $k=1,2,3$.

The asymptotic conditional bias and variance of the circular regression estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$, for $p=0,1$, given in (3.14), can be derived by using the asymptotic bias and variance of estimators $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$, $j=1,2$, as well as the asymptotic covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};p)$, for $p=0,1$. The asymptotic bias of estimators $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$, $j=1,2$ coincides with those obtained for independent data and they were given in (3.16) and (3.23), for $p=0$ and $p=1$, respectively. The asymptotic variance of estimators $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$, $j=1,2$, for $p=0,1$, and the asymptotic covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};p)$, for $p=0,1$ are given below.

It should be noted that although assumption (C2) establishes conditions about correlations $\rho_{k,n}$, $k=1,2,3$, directly derived from model (4.5), using the sine and cosine addition formulas, it is straightforward to obtain equations (4.8)–(4.10), which

relate these covariances with those coming from models (4.6) and (4.7).

Lemma 4.1 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *Under assumptions* (A1), (C1), (C2), (H2), (K1) and (K2), *if* $\mathbf{x}$ *is an interior point of the support of* f, *then, for* $j = 1, 2$, *and* $p = 0, 1$,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\hat{m}_{j,\mathbf{H}}(\mathbf{x}; p) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K)[s_j^2(\mathbf{x}) + f(\mathbf{x})C_j(\mathbf{x})]}{n|\mathbf{H}|f(\mathbf{x})} \\
&\quad + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \\
\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; p), \hat{m}_{2,\mathbf{H}}(\mathbf{x}; p) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] &= \frac{R(K)[c(\mathbf{x}) + f(\mathbf{x})C_3(\mathbf{x})]}{n|\mathbf{H}|f(\mathbf{x})} \\
&\quad + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right),
\end{aligned}$$

where

$$C_1(\mathbf{x}) = \sigma_2^2 f_1^2(\mathbf{x})\rho_{c_2} + 2\sigma_{12} f_1(\mathbf{x}) f_2(\mathbf{x})\rho_{c_3} + \sigma_1^2 f_2^2(\mathbf{x})\rho_{c_1}, \tag{4.11}$$

$$C_2(\mathbf{x}) = \sigma_2^2 f_2^2(\mathbf{x})\rho_{c_2} - 2\sigma_{12} f_1(\mathbf{x}) f_2(\mathbf{x})\rho_{c_3} + \sigma_1^2 f_1^2(\mathbf{x})\rho_{c_1}, \tag{4.12}$$

$$C_3(\mathbf{x}) = \sigma_2^2 f_1(\mathbf{x}) f_2(\mathbf{x})\rho_{c_2} - \sigma_{12} f_1^2(\mathbf{x})\rho_{c_3} + \sigma_{12} f_2^2(\mathbf{x})\rho_{c_3} - \sigma_1^2 f_1(\mathbf{x}) f_2(\mathbf{x})\rho_{c_1}. \tag{4.13}$$

Proof See Section 4.7.

Note that the asymptotic conditional bias of estimators $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; p)$, for $j = 1, 2$, and $p = 0, 1$, considering models (4.6) and (4.7), are the same as those obtained in Lemmas 3.1 and 3.2, for independent data. This is also the case for the asymptotic conditional bias of $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)$, given in (3.17) and (3.24), for $p = 0$ and $p = 1$, respectively. However, the asymptotic conditional variance of $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)$ depends on the spatial correlation. Considering an interior point in the support of the design density f, the Nadaraya–Watson- and local linear-type estimators of m^c have the same asymptotic conditional variance. The following theorem provides this result.

Theorem 4.1 *Let* $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ *be a random sample from a density supported on* $\mathcal{D} \times \mathbb{T}$. *Then, under assumptions* (A1), (C1), (C2), (H2), (K1) and (K2), *the asymptotic conditional variance of* $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)$, *for* $p = 0, 1$, *at a fixed interior point* $\mathbf{x}$ *in the*

support of f, is given by:

$$\mathbb{V}\mathrm{ar}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{R(K)\sigma_1^2\{1+f(\mathbf{x})\rho_{c_1}\}}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} + o_p\left(\frac{1}{n|\mathbf{H}|}\right). \quad (4.14)$$

Proof See Section 4.7.

Notice that the expressions of the asymptotic conditional variance of estimators $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$, for $p=0,1$, have a similar structure to those obtained for the Nadaraya–Watson and local linear estimators in a regression model with Euclidean response and spatially correlated errors. For independent data, it follows that $\rho_{c_1}=0$ in Theorem 4.1 and, consequently, the asymptotic conditional variance of both estimators coincides with the expressions obtained for independent data in (3.18) or (3.25), for $p=0$ and $p=1$, respectively.

The AMSE of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$, defined as the sum of the square of the leading term of the bias (3.17) and the leading term of the variance (4.14), is given in this framework by:

$$\begin{aligned}
&\mathrm{AMSE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)] \\
&= \left\{\frac{1}{2}\mu_2(K)\mathrm{tr}[\mathbf{H}^2\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})] + \frac{\mu_2(K)}{\ell(\mathbf{x})f(\mathbf{x})}\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})\mathbf{H}^2\boldsymbol{\nabla}(\ell f)(\mathbf{x})\right\}^2 \\
&\quad + \frac{R(K)\sigma_1^2\{1+f(\mathbf{x})\rho_{c_1}\}}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})} \\
&= \frac{1}{4}\mu_2^2(K)\mathrm{tr}^2\left(\mathbf{H}^2\left\{\frac{1}{\ell(\mathbf{x})f(\mathbf{x})}[\boldsymbol{\nabla}(\ell f)(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x}) + \boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}(\ell f)(\mathbf{x})] + \boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})\right\}\right) \\
&\quad + \frac{R(K)\sigma_1^2\{1+f(\mathbf{x})\rho_{c_1}\}}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}. \qquad (4.15)
\end{aligned}$$

An asymptotically optimal local bandwidth matrix, $\mathbf{H}^{c,s}_{\mathrm{opt}}(\mathbf{x};0)$, for $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$ can be directly derived minimizing equation (4.15), with respect to $\mathbf{H}$:

$$\mathbf{H}^{c,s}_{\mathrm{opt}}(\mathbf{x};0) = \left\{\frac{R(K)|\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x})|^{1/2}\sigma_1^2[1+f(\mathbf{x})\rho_{c_1}]}{nd\mu_2^2(K)\ell^2(\mathbf{x})f(\mathbf{x})}\right\}^{1/(d+4)} \cdot [\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x})]^{-1/2}, \quad (4.16)$$

where $\tilde{\boldsymbol{\mathcal{G}}}^c(\mathbf{x})$ was given in (3.21).

Similarly, an optimal local bandwidth can be also obtained for the local linear type estimator. In this case, the AMSE of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ is given by:

$$\begin{aligned}&\mathrm{AMSE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)]\\ =\;&\frac{1}{4}\mu_2^2(K)\mathrm{tr}^2\left(\mathbf{H}^2\left\{\frac{1}{\ell(\mathbf{x})}[\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})+\boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}\ell(\mathbf{x})]+\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})\right\}\right)\\ &+\frac{R(K)\sigma_1^2\{1+f(\mathbf{x})\rho_{c_1}\}}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}.\end{aligned}$$

Consequently, the bandwidth matrix which minimizes this expression coincides with (4.16), but taking $\boldsymbol{\mathcal{G}}^c(\mathbf{x}) = \ell^{-1}(\mathbf{x})[\boldsymbol{\nabla}\ell(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}m^c(\mathbf{x})+\boldsymbol{\nabla}m^c(\mathbf{x})\boldsymbol{\nabla}^{\mathrm{T}}\ell(\mathbf{x})]+\boldsymbol{\mathcal{H}}_{m^c}(\mathbf{x})$.

Local bandwidth matrices may be useful for estimating the trend at a given point $\mathbf{x}\in\mathcal{D}\subset\mathbb{R}^d$, however, the nonparametric estimators computed with them may not be accurate enough for reconstructing the whole trend. To get an asymptotic global optimal bandwidth matrix, $\mathbf{H}$ can be selected minimizing the AMISE. Unfortunately, as it was pointed out in Chapter 3 for independent data, there is not a closed form solution for this optimization problem. Moreover, the optimal bandwidth matrices, depending on unknown quantities, cannot be used for practical purposes. Practical bandwidth selection rules, based on cross-validation methods, are considered in what follows.

4.4.1 Bandwidth matrix selection

This section presents cross-validation criteria to select the bandwidth matrix for computing $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$, $p=0,1$, in practice. The CV^c criterion given in (3.33), as well as other smoothing parameter selection methods in nonparametric regression, should not be directly used for selecting the bandwidth when working with dependent data, given that its expectation is severely affected by the correlation (Liu, 2001; Opsomer et al., 2001). In the present setting, the CV^c criterion should be modified in order to account for the effect of the spatial correlation. With this issue in mind, we propose a modified cross-validation (MCV^c) criterion, which selects the bandwidth matrix $\mathbf{H}$

that minimizes the function:

$$\mathrm{MCV}^c(\mathbf{H}) = \sum_{i=1}^{n} \left\{1 - \cos\left[\Theta_i - \hat{m}^c_{\mathbf{H},-N_i}(\mathbf{X}_i; p)\right]\right\},$$

where $\hat{m}^c_{\mathbf{H},-N_i}(\mathbf{X}_i; p)$ denotes the Nadaraya–Watson-type estimator ($p = 0$) or the local linear-type estimator ($p = 1$), computed using all observations except those located within a neighborhood of $\mathbf{X}_i$, namely N_i, and evaluated at $\mathbf{X}_i$. For applying this criterion, the size of the neighborhood N_i must be selected. For simplicity, we consider the MCV^c criterion when $N_i = \{\mathbf{X}_j : \|\mathbf{X}_j - \mathbf{X}_i\| \leq l\}$. For $d = 2$, the neighborhood N_i consists of observations within the circle centered at $\mathbf{X}_i$ and radius l. If there is a strong spatial correlation, more observations should be omitted in the bandwidth selection procedure, and consequently, the value of l for constructing N_i should be larger. The use of the CV^c and MCV^c criteria to select the bandwidth matrix is explored in the following section.

4.5 Simulation study

The performance of the proposed estimators and the cross-validation bandwidth selection criteria are analyzed in a simulation study for $d = 2$. Considering regression model (4.5), 500 samples of size n ($n = 100, 225, 400$) are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$. The regression functions in models M_1 and M_2 defined in Section 3.3.3 will be also considered in this simulation study. Notice that in this case, $\mathbf{X} = (X_1, X_2)$ are spatial locations.

Two different procedures are used to generate the circular spatially correlated errors in model (4.5): the wrapped (Jona-Lasinio et al., 2012) and projected Gaussian approaches (Wang and Gelfand, 2014), introduced in Section 4.2. For each sample, Nadaraya–Watson- and local linear-type estimators of the circular regression function given in (3.14) are computed. In both cases, a multiplicative triweight kernel is considered, while the smoothing matrix $\mathbf{H}$ is chosen by using CV^c and MCV^c criteria. Different values of the radius l are considered in the MCV^c method. Given that the covariates are located in unit square, we set $l(b) = \sqrt{2}b/10$, where $b = 0$ would correspond with the CV^c method and $b = 10$ would provide the maximum

distance between two points in the unit square. After some tests, only three values of b ($b = 1, 2, 3$) are considered. The corresponding CV^c and MCV^c bandwidths are denoted by $\mathbf{H}_{\mathrm{CV}^c}$ and $\mathbf{H}_{\mathrm{MCV}_b^c}$, $b = 1, 2, 3$, respectively. Taking into account the structure of the regression functions, and in order to speed up the computing times, the bandwidth matrix is restricted to be diagonal with possibly different elements. The CASE given in (3.34) is used as a comparative error measure (Kim and SenGupta, 2017). Additionally, the diagonal optimal bandwidth matrix $\mathbf{H}_{\mathrm{CASE}}$ minimizing (3.34), obtained by intensive search, is also computed.

In the first part of the simulation study, the circular spatial errors are generated using the wrapping approach described in Section 4.2.1. In this case, to obtain a realization of the circular (error) process, $\{\varepsilon_i,\ i = 1, \ldots, n\}$, a real-valued Gaussian spatial process Y_i, $i = 1, \ldots, n$, following the model given in (4.1), with zero mean and exponential covariance function (4.2) is considered. The value of the variance σ^2 in (4.2) is fixed equal to one and different values of parameter a_e are considered: $a_e = 0.1$ (weak correlation), $a_e = 0.3$ (medium correlation) and $a_e = 0.6$ (strong correlation). Then, a realization of a wrapped spatial process $\{\varepsilon_i,\ i = 1, \ldots, n\}$ is obtained by:

$$\varepsilon_i = Y_i (\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n.$$

Notice that although the vector of circular variables $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ has almost zero mean direction, to properly apply the estimation procedure in practice, $\boldsymbol{\varepsilon}$ must be centered.

Considering the regression function of model M_1, Table 4.1 shows the average, over 500 replicates, of the CASE given in (3.34) considering the bandwidths selected by the CV^c and MCV_l^c methods, and the minimum value of $\mathrm{CASE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)]$, which can be viewed as a benchmark. Note that the optimal error increases as the dependence range becomes larger. It should be noted first the poor behavior of the CV^c bandwidth, as expected, providing average values of the CASE far from the optimal value, not even decreasing for large sample sizes. In general, MCV^c criterion appears to provide a significant improvement over CV^c when correlation is present. It can be observed that $\mathbf{H}_{\mathrm{MCV}_3^c}$ provides good results for all cases, decreasing the error as n gets larger. For stronger dependence (larger range values), this is the only

Estimator	a_e	n	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{MCV}_1^c}$	$\mathbf{H}_{\mathrm{MCV}_2^c}$	$\mathbf{H}_{\mathrm{MCV}_3^c}$	$\mathbf{H}_{\mathrm{CASE}}$
Nadaraya–Watson	0.1	100	0.2087	0.0902	0.0721	0.0609	0.0387
		225	0.2880	0.1291	0.0768	0.0602	0.0365
		400	0.2932	0.1195	0.0702	0.0585	0.0359
	0.3	100	0.2852	0.1752	0.1342	0.0803	0.0529
		225	0.3080	0.2054	0.1500	0.0788	0.0520
		400	0.2764	0.1967	0.1351	0.0778	0.0497
	0.6	100	0.2316	0.1591	0.1316	0.0806	0.0677
		225	0.2417	0.1775	0.1455	0.0798	0.0620
		400	0.2177	0.1701	0.1325	0.0778	0.0569
Local Linear	0.1	100	0.1818	0.0933	0.0839	0.0672	0.0550
		225	0.2649	0.1190	0.0782	0.0667	0.0532
		400	0.2920	0.1114	0.0771	0.0667	0.0518
	0.3	100	0.2499	0.1651	0.1474	0.1143	0.1062
		225	0.2979	0.2026	0.1546	0.1196	0.1053
		400	0.2785	0.1965	0.1495	0.1168	0.1019
	0.6	100	0.2117	0.1520	0.1392	0.1146	0.1097
		225	0.2361	0.1783	0.1488	0.1212	0.1093
		400	0.2192	0.1725	0.1413	0.1171	0.1074

Table 4.1: Results obtained when the errors in model (4.5) are simulated from wrapped Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (3.34), for the regression function of model M_1, using Nadaraya–Watson- and local linear-type estimators (left and right, respectively). Bandwidth matrix is selected by minimizing $\mathrm{CV}^c(\mathbf{H}_{\mathrm{CV}^c})$, $\mathrm{MCV}^c(\mathbf{H}_{\mathrm{MCV}_1^c}, \mathbf{H}_{\mathrm{MCV}_2^c}, \mathbf{H}_{\mathrm{MCV}_3^c})$ and CASE ($\mathbf{H}_{\mathrm{CASE}}$) as a benchmark.

selector that provides a reasonable behavior. Similar conclusions can be deduced when considering the regression function of model M_2. The corresponding results are displayed in Table 4.2. Notice that when $a_e = 0.1$ (weak spatial correlation), the best behavior is observed when $\mathbf{H}_{\mathrm{MCV}_2^c}$ is employed. As expected, for larger values of the practical range a_e, $\mathbf{H}_{\mathrm{MCV}_3^c}$ provides better results. Note that no major differences have been found if the Nadaraya–Watson or local linear type estimators are employed.

In the second part of the simulation study, the circular spatial errors are generated using the projected approach described in Section 4.2.2. In this case, to obtain a realization of a circular spatial (error) process $\{\varepsilon_i,\ i = 1, \ldots, n\}$, the observations $\mathbf{Y}_i = (Y_{1i}, Y_{2i})$ in model (4.3) are generated setting $\boldsymbol{\mu} = (1, 1)^{\mathrm{T}}$ in (4.3) (to

Estimator	a_e	n	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{MCV}_1^c}$	$\mathbf{H}_{\mathrm{MCV}_2^c}$	$\mathbf{H}_{\mathrm{MCV}_3^c}$	$\mathbf{H}_{\mathrm{CASE}}$
Nadaraya–Watson	0.1	100	0.2231	0.1331	0.1233	0.1282	0.0871
		225	0.3044	0.1475	0.1120	0.1122	0.0792
		400	0.2941	0.1369	0.1008	0.1051	0.0748
	0.3	100	0.1927	0.1740	0.1622	0.1531	0.1277
		225	0.2053	0.1879	0.1620	0.1519	0.1030
		400	0.2154	0.1936	0.1529	0.1414	0.1017
	0.6	100	0.2392	0.1704	0.1605	0.1561	0.1440
		225	0.1891	0.1765	0.1729	0.1553	0.1196
		400	0.1947	0.1687	0.1624	0.1467	0.1063
Local Linear	0.1	100	0.1804	0.1163	0.1092	0.1114	0.0781
		225	0.2574	0.1264	0.0976	0.0980	0.0704
		400	0.2916	0.1160	0.0886	0.0973	0.0668
	0.3	100	0.1991	0.1688	0.1627	0.1526	0.1299
		225	0.2090	0.1869	0.1563	0.1518	0.1277
		400	0.2189	0.1875	0.1505	0.1432	0.1208
	0.6	100	0.2072	0.1605	0.1602	0.1597	0.1348
		225	0.1903	0.1717	0.1669	0.1594	0.1343
		400	0.1962	0.1653	0.1608	0.1509	0.1227

Table 4.2: Results obtained when the errors in model (4.5) are simulated from wrapped Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (3.34), for the regression function of model M_2, using Nadaraya–Watson- and local linear-type estimators (left and right, respectively). Bandwidth matrix is selected by minimizing $\mathrm{CV}^c(\mathbf{H}_{\mathrm{CV}^c})$, $\mathrm{MCV}^c(\mathbf{H}_{\mathrm{MCV}_1^c}, \mathbf{H}_{\mathrm{MCV}_2^c}, \mathbf{H}_{\mathrm{MCV}_3^c})$ and CASE ($\mathbf{H}_{\mathrm{CASE}}$) as a benchmark.

ensure unimodality of the errors and thus obtain *homogeneous* samples) and cross-covariance function given in (4.4) with $\sigma = 1$ and $\tau = 0.9$, to better convey the dependence structure from the linear to the circular process (see Wang and Gelfand, 2014, Figure 4). Different degrees of spatial dependence were chosen in expression (4.4), considering values of $a_e = 0.1$ (weak correlation), $a_e = 0.3$ (medium correlation) and $a_e = 0.6$ (strong correlation). Finally, a realization of the circular spatial process $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}$ is obtained by:

$$\varepsilon_i = \operatorname{atan2}(Y_{2i}, Y_{1i}), \quad i = 1, \ldots, n.$$

Notice that as for the wrapping approach, the realization of the circular error

Estimator	a_e	n	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{MCV}_1^c}$	$\mathbf{H}_{\mathrm{MCV}_2^c}$	$\mathbf{H}_{\mathrm{MCV}_3^c}$	$\mathbf{H}_{\mathrm{CASE}}$
Nadaraya–Watson	0.1	100	0.1465	0.0629	0.0482	0.0452	0.0249
		225	0.2207	0.0874	0.0417	0.0386	0.0216
		400	0.2639	0.0722	0.0380	0.0385	0.0209
	0.3	100	0.2562	0.1746	0.1382	0.1142	0.0564
		225	0.2235	0.1851	0.1248	0.1123	0.0558
		400	0.2498	0.1987	0.1279	0.1109	0.0474
	0.6	100	0.2368	0.1871	0.1567	0.1378	0.0620
		225	0.2452	0.2126	0.1659	0.1371	0.0585
		400	0.2424	0.1991	0.1600	0.1303	0.0525
Local Linear	0.1	100	0.1376	0.0735	0.0667	0.0596	0.0381
		225	0.1976	0.0891	0.0552	0.0566	0.0321
		400	0.2430	0.0745	0.0497	0.0535	0.0301
	0.3	100	0.2452	0.1823	0.1658	0.1534	0.1205
		225	0.2300	0.1948	0.1555	0.1498	0.1203
		400	0.2524	0.1974	0.1554	0.1478	0.1125
	0.6	100	0.1988	0.1772	0.1703	0.1622	0.1376
		225	0.2072	0.1927	0.1710	0.1612	0.1361
		400	0.2051	0.1868	0.1637	0.1578	0.1300

Table 4.3: Results obtained when the errors in model (4.5) are simulated from projected Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (3.34), for the regression function of model M_1, using Nadaraya–Watson- and local linear-type estimators (left and right, respectively). Bandwidth matrix is selected by minimizing $\mathrm{CV}^c(\mathbf{H}_{\mathrm{CV}^c})$, $\mathrm{MCV}^c(\mathbf{H}_{\mathrm{MCV}_1^c}, \mathbf{H}_{\mathrm{MCV}_2^c}, \mathbf{H}_{\mathrm{MCV}_3^c})$ and CASE ($\mathbf{H}_{\mathrm{CASE}}$) as a benchmark.

process $\boldsymbol{\varepsilon}$ must be centered.

Considering the regression function of model M_1, numerical results are summarized in Table 4.3. For this regression function, as it was pointed in Section 4.2.1 for wrapped Gaussian spatial processes, when the CV^c bandwidth is used, the CASE corresponding to the CV^c bandwidth matrix is the largest in all the scenarios. Regarding the MCV^c criterion, when the dependence structure is stronger, the value of b must be larger (that is, the value of the radius l must be larger). For example, considering a weak dependence structure ($a_e = 0.1$), the use of $\mathbf{H}_{\mathrm{MCV}_2^c}$ seems to show a slightly better performance. If the dependence structure is stronger, $\mathbf{H}_{\mathrm{MCV}_3^c}$ provides better results. Table 4.4 shows the results for the regression function M_2.

Numerical outputs are completed with some additional plots. Given that similar

Estimator	a_e	n	$\mathbf{H}_{\mathrm{CV}^c}$	$\mathbf{H}_{\mathrm{MCV}_1^c}$	$\mathbf{H}_{\mathrm{MCV}_2^c}$	$\mathbf{H}_{\mathrm{MCV}_3^c}$	$\mathbf{H}_{\mathrm{CASE}}$
Nadaraya–Watson	0.1	100	0.1764	0.1111	0.1055	0.1083	0.0813
		225	0.2324	0.1177	0.0865	0.0901	0.0578
		400	0.2689	0.1048	0.0772	0.0854	0.0516
	0.3	100	0.2701	0.2132	0.1822	0.1752	0.1457
		225	0.2890	0.2354	0.1919	0.1723	0.1043
		400	0.2850	0.2223	0.1734	0.1583	0.0940
	0.6	100	0.2458	0.2068	0.1934	0.1847	0.1472
		225	0.2558	0.2211	0.1953	0.1777	0.1010
		400	0.2444	0.2082	0.1806	0.1679	0.0960
Local Linear	0.1	100	0.1551	0.1023	0.0935	0.0957	0.0792
		225	0.2023	0.1024	0.0769	0.0790	0.0513
		400	0.2470	0.0909	0.0701	0.0717	0.0465
	0.3	100	0.2510	0.1998	0.1832	0.1796	0.1621
		225	0.2708	0.2268	0.1913	0.1749	0.1412
		400	0.2773	0.2134	0.1769	0.1687	0.1326
	0.6	100	0.2344	0.1995	0.1900	0.1877	0.1849
		225	0.2441	0.2143	0.1923	0.1856	0.1517
		400	0.2403	0.2017	0.1801	0.1752	0.1439

Table 4.4: Results obtained when the errors in model (4.5) are simulated from projected Gaussian spatial processes. Average (over 500 replicates) of the CASE given in (3.34), for the regression function of model M_2, using Nadaraya–Watson- and local linear-type estimators (left and right, respectively). Bandwidth matrix is selected by minimizing $\mathrm{CV}^c(\mathbf{H}_{\mathrm{CV}^c})$, $\mathrm{MCV}^c(\mathbf{H}_{\mathrm{MCV}_1^c}, \mathbf{H}_{\mathrm{MCV}_2^c}, \mathbf{H}_{\mathrm{MCV}_3^c})$ and CASE ($\mathbf{H}_{\mathrm{CASE}}$) as a benchmark.

results were obtained for $\hat{m}^c_{\mathbf{H}}(\mathbf{x};0)$ and $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ in the previous simulations, plots are only shown for $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$. As an illustration of the appropriate performance of the estimator $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$, Figure 4.4 shows the theoretical regression functions for models M_1 and M_2 (left panels) and the corresponding average, over 500 replicates, of the fit values using $\hat{m}^c_{\mathbf{H}}(\mathbf{x};1)$ considering samples of size $n = 400$ and when the circular errors are generated from a wrapped Gaussian spatial process (center panels) and from a projected Gaussian spatial process (right panels). In this example, for both types of circular errors, an exponential covariance model is used with range parameter equal to 0.3. Estimates are computed employing the bandwidth matrix $\mathbf{H}_{\mathrm{MCV}_3^c}$. Notice that, for comparison purposes, the theoretical regression functions are plotted in a 30×30 regular grid on the covariate region (the same grid where

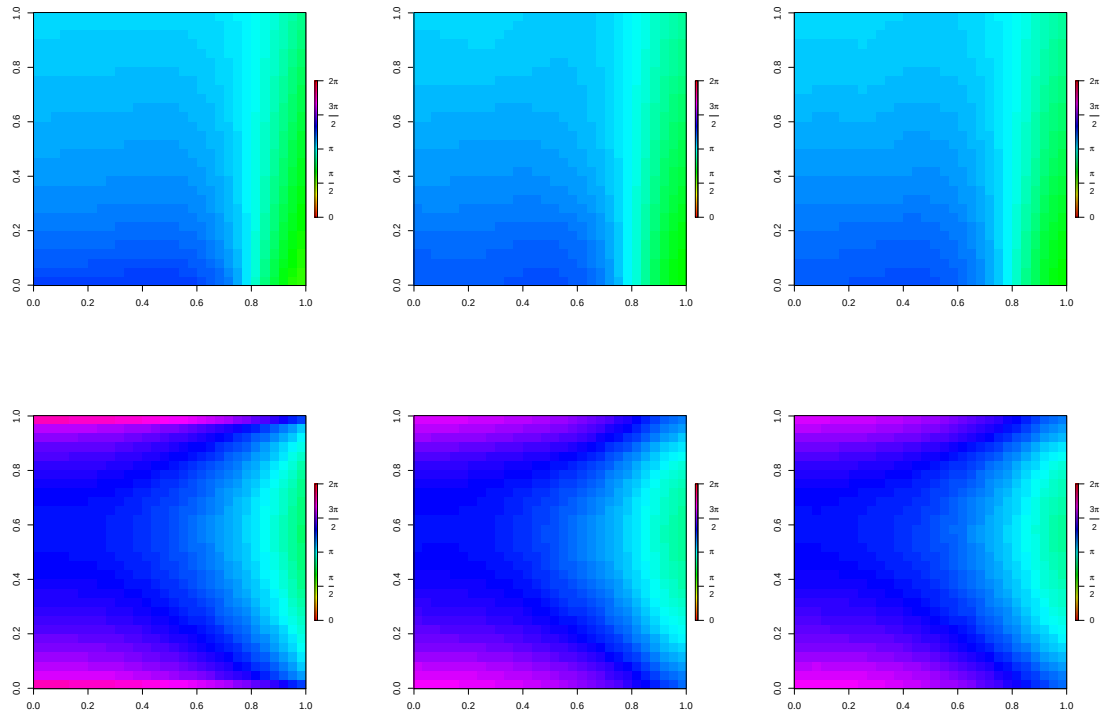

Figure 4.4: Theoretical regression function (left panels), jointly with the local linear-type estimators using wrapped (center panels) and projected Gaussian spatial processes (right panels) to generate the errors in model (4.5), for the regression function of model M_1 (top row panels) and M_2 (bottom row panels).

the estimations are computed). Plots in the top row present the results for the data generated using M_1 and those in the bottom row using M_2. The estimation of the circular trend surfaces seems to be quite accurate, no matter the approach (wrapped or projected) used to generate the circular spatial errors.

4.6 Real data illustration

In this section, to illustrate our nonparametric proposal, an oceoanography dataset, previously analized by several authors, containing wave directions recorded in 1494 grid points on the Adriatic Sea area (from a calm period transitioning to a storm period at different times) is considered. These data outputs were derived from a wave

Figure 4.5: Random sample of 150 wave directions in the Adriatic Sea area on April 2, 2010 at 6am during a calm period.

model implemented by Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA) and they are available in the `R` package `CircSpaceTime` (Jona-Lasinio et al., 2019). In this illustration, we only consider wave directions for a calm period, corresponding to measurements taken at 06:00 on April 2 at Adriatic Sea. Figure 4.5 shows a random sample of size 150 of those observations. As intuition suggests, these data seem to exhibit a spatial pattern.

We consider the linear-circular regression model given in (4.5), where $\mathbf{X}_i = (X_{i1}, X_{i2})$, for $i = 1, \ldots, 1494$, represent the different locations, with X_{i1} the longitude and X_{i2} the latitude, and Θ_i the corresponding wave direction at that location. The nonparametric estimator of the circular regression function, given in (3.14), is computed. Taking into account that the performance of $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 0)$ and $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 1)$ was similar in the simulation study, only results employing $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 1)$ are shown in this application. As for the simulation study, a multiplicative triweight kernel is considered.

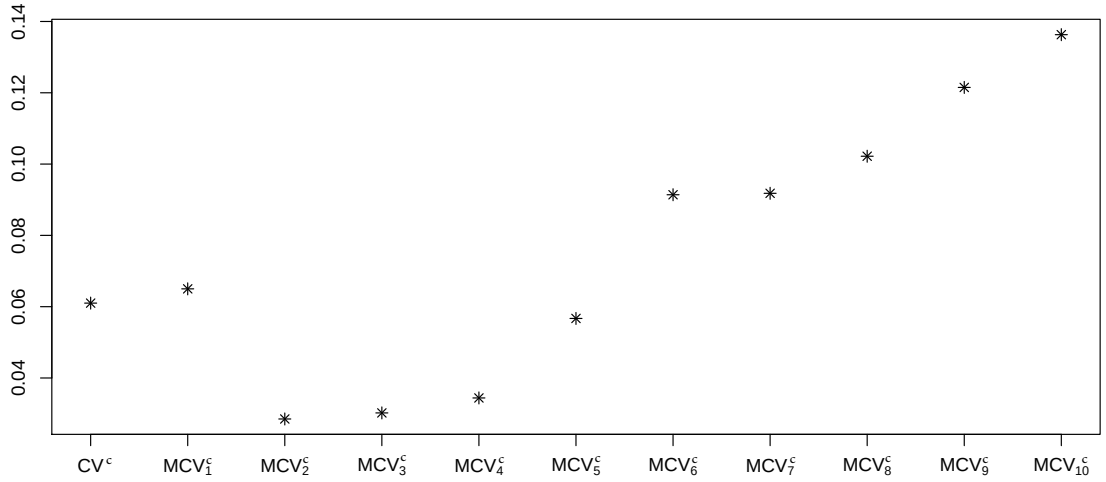

Figure 4.6: Prediction errors given in (4.17) for each bandwidth matrix selected by CV^c and $\mathrm{MCV}_b^c, b = 1, \ldots, 10$.

The bandwidth matrix is selected employing a cross-validation criterion. In order to decide if using CV^c or MCV^c (and, in that case, a suitable value for the radius l), the whole sample is split in two parts, a randomly selected training sample of size 1345 (90% of the data), denoted by $\{(\tilde{\mathbf{X}}_i, \tilde{\Theta}_i)\}_{i=1}^{1345}$, and a test sample, made up of the remaining observations, of size 149 (10% of the data), denoted by $\{(\check{\mathbf{X}}_j, \check{\Theta}_j)\}_{j=1}^{149}$. Then, estimations at each testing point $\check{\mathbf{X}}_j, j = 1, \ldots, 149$, with different bandwidths, are compared with the testing responses using the following prediction error:

$$\sum_{j=1}^{149} \left\{1 - \cos\left[\check{\Theta}_j - \hat{m}_{\hat{\mathbf{H}}}(\check{\mathbf{X}}_j; 1)\right]\right\}, \tag{4.17}$$

where $\hat{m}_{\hat{\mathbf{H}}}(\check{\mathbf{X}}_j; 1)$ is the local linear-type circular regression estimator computed using the training sample and evaluated at the testing point $\check{\mathbf{X}}_j, j = 1, \ldots, 149$, and $\hat{\mathbf{H}}$ denotes the bandwidth matrix selected using CV^c or MCV^c, employing the training sample. In the case of the MCV^c criterion, different values of the radius l are considered. As in the simulation study, we set $l(b) = \sqrt{2}b/10$, now with $b = 1, \ldots, 10$. These bandwidth matrices are searched in the family of the symmetric and definite positive full bandwidth matrices, using an optimization algorithm based on the Nelder–Mead simplex method described in (Lagarias et al., 1998). To apply this opti-

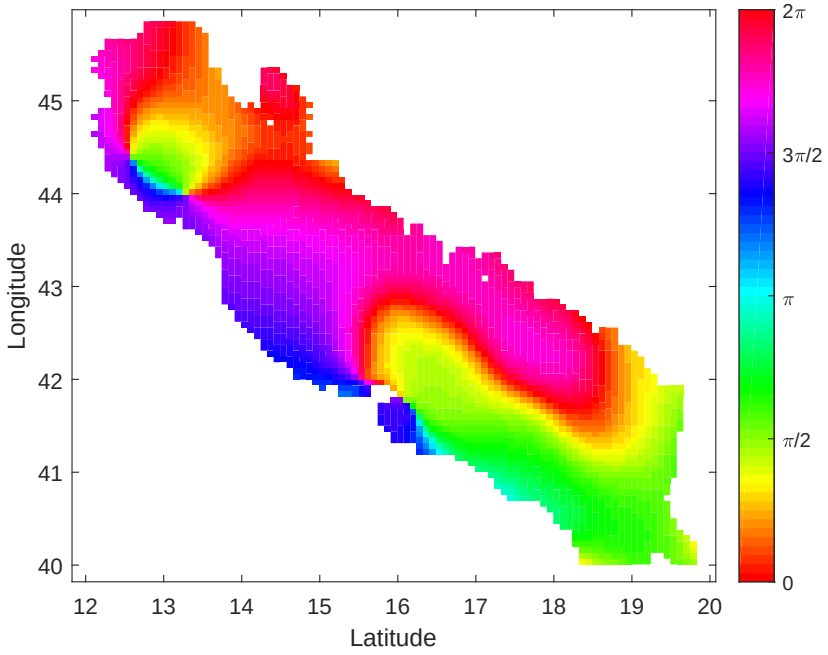

Figure 4.7: Regression function estimation using the local linear-type estimator $\hat{m}^{c}_{\hat{\mathbf{H}}}(\mathbf{x};0)$, using the bandwidth matrix $\mathbf{H}$ given in (4.18), selected with the MCV^c ($b=2$) criterion.

mization procedure, we use the initial bandwidth matrix $\mathbf{H}_{\text{init}} = 1.5{\cdot}\text{diag}\left\{\hat{\sigma}_{\tilde{X}_1}, \hat{\sigma}_{\tilde{X}_2}\right\}$, where $\hat{\sigma}_{\tilde{X}_1}$ and $\hat{\sigma}_{\tilde{X}_2}$, with $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2)$, are the training sample standard deviations of $\tilde{X}_1$ and $\tilde{X}_2$, respectively. Figure 4.6 shows the prediction error given in (4.17) for each bandwidth matrix $\hat{\mathbf{H}}$. It can be seen that the minimum error is achieved when MCV^c, with $b=2$, is employed, converging the algorithm when using this criterion to

$$\mathbf{H}_{\text{MCV}^c_2} = \begin{pmatrix} 0.4744 & 0.0081 \\ 0.0081 & 0.3529 \end{pmatrix}. \tag{4.18}$$

Figure 4.7 shows the circular trend surface estimation using $\mathbf{H}_{\text{MCV}^c_2}$, given in (4.18). The estimation grid was constructed by overlying the survey values of longitude and latitude with a 100×100 grid and, then, dropping every grid point that did not satisfy at least one of the following two requirements: (a) it is within two "grid cell length" from an observation point, or (b) the calculation for the estimates of the sine and cosine components at that grid point uses a smoothing vector that is

sufficiently stable. The sine and cosine of the detrended wave direction dataset were tested for isotropy and stationarity, following the proposals by Bowman and Crujeiras (2013). For both tests, p-values were larger than the usual significance levels (for isotropy: 0.3206 and 0.1271 for sine and cosine, respectively; for stationarity, p-values were larger than 0.99 for both processes).

From Figure 4.7, it can be clearly seen the shoreline orientation of the waves (recall that our measurements correspond to a calm period), providing the different color pattern along the coastline. Something which is interesting to notice is the behavior in the Gulf, where waves rotate to different directions, and a main current can be also observed. According to this pattern, more variation can be observed in the North, something that was also pointed out by Jona-Lasinio et al. (2012), although for a storm period.

4.7 Proofs of the main results

This section is devoted to present the proof of Lemma 4.1 and Theorem 4.1. The asymptotic variance of the proposed nonparametric regression estimator $\hat{m}^{c}_{\mathbf{H}}(\mathbf{x}; p)$, for $p = 0, 1$, given in (3.14), for spatially correlated data, is calculated in Theorem 4.1. Some previous results are given in Lemma 4.1.

Proof of Lemma 4.1 First, the asymptotic variance of $\hat{m}_{j,\mathbf{H}}(\mathbf{x}; p)$ is derived. When $p = 0$, if $j = 1$, it follows that

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x}; 0) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] \\
&= \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})}{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \sum_{j=1}^{n} K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x})} \mathbb{C}\mathrm{ov}[\sin(\Theta_i), \sin(\Theta_j) \mid \mathbf{X}_1, \ldots, \mathbf{X}_n] \\
&= \frac{\sum_{i=1}^{n} K^2_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) s_1^2(\mathbf{X}_i)}{[\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})]^2} + \frac{\sum_{i \neq j} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) K_{\mathbf{H}}(\mathbf{X}_j - \mathbf{x}) C_{n,1}(\mathbf{X}_i, \mathbf{X}_j)}{[\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})]^2}.
\end{aligned}$$

Using (A.4) of Lemma A.1 and (A.7) of Lemma A.2, it is obtained that

$$\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) = f(\mathbf{x}) + o_{\mathbb{P}}(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})s_1^2(\mathbf{X}_i) = \frac{1}{|\mathbf{H}|}R(K)f(\mathbf{x})s_1^2(\mathbf{x}) + o_{\mathbb{P}}(1). \tag{4.19}$$

and

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})C_{n,1}(\mathbf{X}_i,\mathbf{X}_j)\\
&= \frac{1}{n^2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})[f_1(\mathbf{X}_i)f_1(\mathbf{X}_j)\sigma_2^2\rho_{2,n}(\mathbf{X}_i-\mathbf{X}_j)\\
&\quad +f_1(\mathbf{X}_i)f_2(\mathbf{X}_j)\sigma_{12}\rho_{3,n}(\mathbf{X}_i-\mathbf{X}_j) + f_2(\mathbf{X}_i)f_1(\mathbf{X}_j)\sigma_{12}\rho_{3,n}(\mathbf{X}_i-\mathbf{X}_j)\\
&\quad +f_2(\mathbf{X}_i)f_2(\mathbf{X}_j)\sigma_1^2\rho_{1,n}(\mathbf{X}_i-\mathbf{X}_j)]\\
&= \frac{1}{n|\mathbf{H}|}R(K)f(\mathbf{x})[\sigma_2^2 f_1^2(\mathbf{x})\rho_{c_2} + 2\sigma_{12}f_1(\mathbf{x})f_2(\mathbf{x})\rho_{c_3} + \sigma_1^2 f_2^2(\mathbf{x})\rho_{c_1}]\\
&= \frac{1}{n|\mathbf{H}|}R(K)f^2(\mathbf{x})C_1(\mathbf{x}) + o_{\mathbb{P}}(1).
\end{aligned} \tag{4.20}$$

Therefore, using also (A.1) of Lemma A.1, it follows that

$$\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[s_1^2(\mathbf{x}) + f(\mathbf{x})C_1(\mathbf{x})] + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Similarly, when $j=2$, it follows that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[s_2^2(\mathbf{x}) + f(\mathbf{x})C_2(\mathbf{x})]\\
&\quad + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

When $p=1$, if $j=1$, one gets that

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n]\\
&= \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}_1^{c,s}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\mathbf{e}_1,
\end{aligned}$$

where $\boldsymbol{\Sigma}_1^{c,s}$ is the covariance matrix of $\sin(\Theta)$, whose (i,j)-entry is given by $\boldsymbol{\Sigma}_1^{c,s}(i,j) = \mathbb{C}\mathrm{ov}[\sin(\Theta_i),\sin(\Theta_j)\mid \mathbf{X}_i,\mathbf{X}_j]$, $i,j=1,\ldots,n$. Using (A.4), (A.5) and (A.6) of Lemma A.1, (A.7) of Lemma A.2, (A.13) and (A.14) of Lemma A.3, with similar arguments

to those employed in (4.19) and (4.20), and defining

$$\begin{aligned}
s_{11,n}(\mathbf{x}) &= n^{-2}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})s_1^2(\mathbf{X}_i),\\
s_{12,n}(\mathbf{x}) &= n^{-2}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})s_1^2(\mathbf{X}_i),\\
s_{13,n}(\mathbf{x}) &= n^{-2}\sum_{i=1}^{n} K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^{\mathrm{T}}s_1^2(\mathbf{X}_i),\\
c_{11,n}(\mathbf{x}) &= n^{-2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})C_{n,1}(\mathbf{X}_i,\mathbf{X}_j),\\
c_{12,n}(\mathbf{x}) &= n^{-2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_i-\mathbf{x})C_{n,1}(\mathbf{X}_i,\mathbf{X}_j),\\
c_{13,n}(\mathbf{x}) &= n^{-2}\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_j-\mathbf{x})^{\mathrm{T}}C_{n,1}(\mathbf{X}_i,\mathbf{X}_j),
\end{aligned}$$

it can be obtained that

$$\begin{aligned}
s_{11,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}R(K)f(\mathbf{x})s_1^2(\mathbf{x}) + o_{\mathbb{P}}\left(n^{-1}|\mathbf{H}|^{-1}\right),\\
s_{12,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_d),\\
s_{13,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_{d\times d}),\\
c_{11,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}R(K)f^2(\mathbf{x})C_1(\mathbf{x}) + o_{\mathbb{P}}\left(n^{-1}|\mathbf{H}|^{-1}\right),\\
c_{12,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_d),\\
c_{13,n}(\mathbf{x}) &= n^{-1}|\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_{d\times d}).
\end{aligned}$$

Therefore, it follows that

$$\begin{aligned}
n^{-2}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}_1^{c,s}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}} &= \begin{pmatrix} s_{11,n}(\mathbf{x})+c_{11,n}(\mathbf{x}) & s_{12,n}^{\mathrm{T}}(\mathbf{x})+c_{12,n}^{\mathrm{T}}(\mathbf{x}) \\ s_{12,n}(\mathbf{x})+c_{12,n}(\mathbf{x}) & s_{13,n}(\mathbf{x})+c_{13,n}(\mathbf{x}) \end{pmatrix}\\
&= \frac{1}{n|\mathbf{H}|}\begin{pmatrix} R(K)f(\mathbf{x})[s_1^2(\mathbf{x})+f(\mathbf{x})C_1(\mathbf{x})]+o_{\mathbb{P}}(1) & o_{\mathbb{P}}(\mathbf{1}_d^{\mathrm{T}}) \\ o_{\mathbb{P}}(\mathbf{1}_d) & o_{\mathbb{P}}(\mathbf{1}_{d\times d}) \end{pmatrix}.
\end{aligned}$$

Consequently, using (3.48) and the previous expression, by straightforward calculations, one gets

$$\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1)\mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[s_1^2(\mathbf{x})+f(\mathbf{x})C_1(\mathbf{x})] + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Similarly, when $j = 2$, it can be obtained that

$$\mathbb{V}\mathrm{ar}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[s_2^2(\mathbf{x}) + f(\mathbf{x})C_2(\mathbf{x})] + \boldsymbol{o}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Regarding the conditional covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};0)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};0)$, using (3.45) and similar arguments to those employed in (4.20), considering (A.1) of Lemma A.1, it follows that

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};0), \hat{m}_{2,\mathbf{H}}(\mathbf{x};0) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&= \frac{\sum_{i=1}^n \sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})}{\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\sum_{j=1}^n K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})}\mathbb{C}\mathrm{ov}[\sin(\Theta_i),\cos(\Theta_j) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&= \frac{\sum_{i=1}^n K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})c(\mathbf{X}_i)}{[\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})]^2} + \frac{\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})C_{n,3}(\mathbf{X}_i,\mathbf{X}_j)}{[\sum_{i=1}^n K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})]^2} \\
&= \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[c(\mathbf{x}) + f(\mathbf{x})C_3(\mathbf{x})] + \boldsymbol{o}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).
\end{aligned}$$

When $p = 1$, the conditional covariance between $\hat{m}_{1,\mathbf{H}}(\mathbf{x};1)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};1)$ is:

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1), \hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\
&= \mathbf{e}_1^{\mathrm{T}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}^{c,s}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}}(\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1}\mathbf{e}_1,
\end{aligned}$$

where $\boldsymbol{\Sigma}^{c,s}$ is the covariance matrix of $\sin(\Theta)$ and $\cos(\Theta)$, whose (i,j)-entry is $\boldsymbol{\Sigma}^{c,s}(i,j) = \mathbb{C}\mathrm{ov}[\sin(\Theta_i),\cos(\Theta_j) \mid \mathbf{X}_i,\mathbf{X}_j]$, $i,j = 1,\ldots,n$. Using (A.7) of Lemma A.2, (A.13) and (A.14) of Lemma A.3, employing similar arguments to those given in (4.19) and (4.20), and defining

$$\begin{aligned}
c_{31,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})C_{n,3}(\mathbf{X}_i,\mathbf{X}_j), \\
c_{32,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_i-\mathbf{x})C_{n,3}(\mathbf{X}_i,\mathbf{X}_j), \\
c_{33,n}(\mathbf{x}) &= n^{-2}\textstyle\sum_{i\neq j} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_j-\mathbf{x})^{\mathrm{T}}C_{n,3}(\mathbf{X}_i,\mathbf{X}_j),
\end{aligned}$$

it can be obtained that

$$c_{31,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}R(K)f^2(\mathbf{x})C_3(\mathbf{x}) + \boldsymbol{o}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \tag{4.21}$$

$$c_{32,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}\boldsymbol{o}_{\mathbb{P}}(\mathbf{1}_d), \tag{4.22}$$

$$c_{33,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}\boldsymbol{o}_{\mathbb{P}}(\mathbf{1}_{d\times d}). \tag{4.23}$$

Therefore, using (3.45), (3.46), (3.47), (4.21), (4.22) and (4.23), it follows that

$$\begin{aligned} n^{-2}\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\Sigma}^{c,s}\boldsymbol{\mathcal{W}}_{\mathbf{x}}\boldsymbol{\mathcal{X}}_{\mathbf{x}} &= \begin{pmatrix} s_{31,n}(\mathbf{x}) + c_{31,n}(\mathbf{x}) & s_{32,n}^{\mathrm{T}}(\mathbf{x}) + c_{32,n}^{\mathrm{T}}(\mathbf{x}) \\ s_{32,n}(\mathbf{x}) + c_{32,n}(\mathbf{x}) & s_{33,n}(\mathbf{x}) + c_{33,n}(\mathbf{x}) \end{pmatrix} \\ &= \frac{1}{n|\mathbf{H}|}\begin{pmatrix} R(K)f(\mathbf{x})[c(\mathbf{x}) + f(\mathbf{x})C_3(\mathbf{x})] + \boldsymbol{o}_{\mathbb{P}}(1) & \boldsymbol{o}_{\mathbb{P}}(\mathbf{1}_d^{\mathrm{T}}) \\ \boldsymbol{o}_{\mathbb{P}}(\mathbf{1}_d) & \boldsymbol{o}_{\mathbb{P}}(\mathbf{1}_{d\times d}) \end{pmatrix}. \end{aligned}$$

Consequently, using (3.48) and the above expression, by straightforward calculations, one gets

$$\begin{aligned} \mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};1), \hat{m}_{2,\mathbf{H}}(\mathbf{x};1) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{n|\mathbf{H}|f(\mathbf{x})}R(K)[c(\mathbf{x}) + f(\mathbf{x})C_3(\mathbf{x})] \\ &\quad + \boldsymbol{o}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \end{aligned}$$

Proof of Theorem 4.1 To derive the variance of $\hat{m}_{\mathbf{H}}^{c}(\mathbf{x};p)$, for $p = 0, 1$, following the arguments used in the proof of Theorem 3.1 and 3.2, one gets that

$$\begin{aligned} \mathbb{V}\mathrm{ar}[\hat{m}_{\mathbf{H}}^{c}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{m}_{2,\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\ &\quad + \frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{V}\mathrm{ar}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\ &\quad - \frac{2m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}\mathbb{C}\mathrm{ov}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};p), \hat{m}_{2,\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] \\ &\quad + \mathcal{O}\big\{[\hat{m}_{1,\mathbf{H}}(\mathbf{x};p) - m_1(\mathbf{x})]^3\big\} \\ &\quad + \mathcal{O}\big\{[\hat{m}_{2,\mathbf{H}}(\mathbf{x};p) - m_2(\mathbf{x})]^3\big\}. \end{aligned}$$

Using Lemma 4.1, one gets that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[\hat{m}^{c}_{\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] &= \frac{1}{n|\mathbf{H}|}\frac{R(K)}{f(\mathbf{x})}\frac{m_1^2(\mathbf{x})}{\ell^4(\mathbf{x})}[s_2^2(\mathbf{x})+f(\mathbf{x})C_2(\mathbf{x})]\\ &\quad+\frac{1}{n|\mathbf{H}|}\frac{R(K)}{f(\mathbf{x})}\frac{m_2^2(\mathbf{x})}{\ell^4(\mathbf{x})}[s_1^2(\mathbf{x})+f(\mathbf{x})C_1(\mathbf{x})]\\ &\quad-\frac{2}{n|\mathbf{H}|}\frac{R(K)}{f(\mathbf{x})}\frac{m_1(\mathbf{x})m_2(\mathbf{x})}{\ell^4(\mathbf{x})}[c(\mathbf{x})+f(\mathbf{x})C_3(\mathbf{x})]\\ &\quad+o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

Using (3.11), (3.12), (3.13), (4.11), (4.12) and (4.13), it follows that

$$\begin{aligned}&m_1^2(\mathbf{x})[s_2^2(\mathbf{x})+f(\mathbf{x})C_2(\mathbf{x})]+m_2^2(\mathbf{x})[s_1^2(\mathbf{x})+f(\mathbf{x})C_1(\mathbf{x})]\\ &\quad-2m_1(\mathbf{x})m_2(\mathbf{x})[c(\mathbf{x})+f(\mathbf{x})C_3(\mathbf{x})]\\ &= f_1^2(\mathbf{x})\ell^2(\mathbf{x})[f_2^2(\mathbf{x})\sigma_2^2-2f_2(\mathbf{x})f_1(\mathbf{x})\sigma_{12}+f_1^2(\mathbf{x})\sigma_1^2+f(\mathbf{x})\sigma_2^2f_2^2(\mathbf{x})\rho_{c_2}\\ &\quad-2f(\mathbf{x})\sigma_{12}f_1(\mathbf{x})f_2(\mathbf{x})\rho_{c_3}+f(\mathbf{x})\sigma_1^2f_1^2(\mathbf{x})\rho_{c_1}]+f_2^2(\mathbf{x})\ell^2(\mathbf{x})[f_1^2(\mathbf{x})\sigma_2^2\\ &\quad+2f_1(\mathbf{x})f_2(\mathbf{x})\sigma_{12}+f_2^2(\mathbf{x})\sigma_1^2+f(\mathbf{x})\sigma_2^2f_1^2(\mathbf{x})\rho_{c_2}+2f(\mathbf{x})\sigma_{12}f_1(\mathbf{x})f_2(\mathbf{x})\rho_{c_3}\\ &\quad+f(\mathbf{x})\sigma_1^2f_2^2(\mathbf{x})\rho_{c_1}]-2f_1(\mathbf{x})f_2(\mathbf{x})\ell^2(\mathbf{x})[f_1(\mathbf{x})f_2(\mathbf{x})\sigma_2^2-f_1^2(\mathbf{x})\sigma_{12}\\ &\quad+f_2^2(\mathbf{x})\sigma_{12}-f_1(\mathbf{x})f_2(\mathbf{x})\sigma_1^2+f(\mathbf{x})\sigma_2^2f_1(\mathbf{x})f_2(\mathbf{x})\rho_{c_2}\\ &\quad-f(\mathbf{x})\sigma_{12}f_1^2(\mathbf{x})\rho_{c_3}+f(\mathbf{x})\sigma_{12}f_2^2(\mathbf{x})\rho_{c_3}-\sigma_1^2f_1(\mathbf{x})f_2(\mathbf{x})\rho_{c_1}]\\ &= \ell^2(\mathbf{x})\sigma_1^2[1+f(\mathbf{x})\rho_{c_1}].\end{aligned}$$

Consequently, it can be directly obtained that

$$\mathbb{V}\mathrm{ar}[\hat{m}^{c}_{\mathbf{H}}(\mathbf{x};p) \mid \mathbf{X}_1,\ldots,\mathbf{X}_n] = \frac{R(K)\sigma_1^2[1+f(\mathbf{x})\rho_{c_1}]}{n|\mathbf{H}|\ell^2(\mathbf{x})f(\mathbf{x})}+o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).$$

Chapter 5

Some ideas on testing parametric regression models with a circular response and an $\mathbb{R}^d$-valued covariate

5.1 Introduction

In a variety of contexts, circular measurements are accompanied by observations of other Euclidean random variables. The joint behavior of these circular and Euclidean variables can be analyzed by considering a regression model, allowing at the same time to explain the possible relation between the variables and, at the same time, also to make predictions on the variable of interest. As pointed out in Chapter 3 of this dissertation, parametric regression estimators for linear-circular models (regression models with a circular response and an $\mathbb{R}^d$-valued covariate) with independent data were studied by Fisher and Lee (1992), Presnell et al. (1998) and Kim and SenGupta (2017), among others. In the presence of spatial correlation (as it was indicated in Chapter 4), for instance, Jona-Lasinio et al. (2012), Wang and Gelfand (2014), Lagona et al. (2015) and Mastrantonio et al. (2016), employed parametric methods to model circular spatial processes. Alternatively, nonparametric kernel-type estimators of the regression function considering a model with a circular

response and a univariate Euclidean covariate were introduced in Di Marzio et al. (2013). The extension to a model with an $\mathbb{R}^d$-valued covariate was considered in Chapter 3 (for independent data) and in Chapter 4 (for spatially correlated data). As pointed out in both chapters, to compute these kernel-type estimators it is crucial to select a bandwidth (a symmetric $d \times d$ matrix for an $\mathbb{R}^d$-valued covariate) which directly impacts the smoothness of the estimator. If the bandwidth matrix is appropriately chosen, these nonparametric methods provide more flexible and robust estimators than those obtained when using parametric approaches, avoiding misspecification problems. However, if a suitable parametric regression model is assumed, parametric methods usually provide estimators which are more efficient and easier to interpret.

At this point, an important question in this context is to decide if a certain parametric family is appropriate to model the unknown circular regression function. If this assumption holds, a parametric method should be preferably used to estimate it. If not, it would possibly be more convenient to use a nonparametric approach to estimate this function. Both approaches, parametric and nonparametric, have been used to analyze different datasets in the literature. For instance, the classical blue periwinkles dataset (which collects measurements of direction and distance moved by 31 blue periwinkles) was analyzed using parametric methods in Fisher and Lee (1992) and Presnell et al. (1998), considering the direction as the response variable and the distance as the covariable. On the other hand, also considering this dataset and this regression model, nonparametric techniques were employed in Di Marzio et al. (2013) to estimate the corresponding regression function. Another such example is the sand hopper orientation dataset (described in Chapter 3 of this dissertation), which was studied by Scapini et al. (2002) using parametric methods. Following the proposal in Presnell et al. (1998), these authors considered a projected multivariate linear model (PMLM) to analyze the orientation of two species of sand hoppers as a function of different covariates. On the other hand, this dataset was also explored using nonparametric tools in Section 3.3.4, considering a regression model with a circular response (sand hopper orientation) and two real-valued covariates (temperature and humidity). In this case, the regression function was estimated nonparametrically using a local linear-type estimator. In order to determine if a parametric regres-

sion model is a suitable representation of such datasets, goodness-of-fit tests can be designed and analyzed, providing a tool for assessing a general class of parametric linear-circular regression models.

There is a substantial literature on testing parametric regression models involving Euclidean data, for example Kozek (1991), Härdle and Mammen (1993), González Manteiga and Vilar Fernández (1995), Biedermann and Dette (2000) and Park et al. (2015), among others. See also González-Manteiga and Crujeiras (2013) for a review on this topic. The previous testing procedures, as well as the one formulated in Chapter 2 of this dissertation, are based on measuring differences between a suitable parametric estimator under the null hypothesis and a nonparametric one. Specifically, L_2-norm or supremum-norm tests, among others, can be employed for testing parametric regression models with Euclidean responses and covariates. In the context of regression models with directional response and directional or Euclidean explanatory variables, the literature on goodness-of-fit tests is relatively scarce. In this setting, in Deschepper et al. (2008), an exploratory tool and a lack-of-fit test for circular-linear regression models (Euclidean response and circular covariates) were proposed. The same problem was studied by García-Portugués et al. (2016), using nonparametric methods. The authors proposed a testing procedure based on the weighted squared distance between a nonparametric and a parametric regression estimator, where the nonparametric regression estimator was obtained by a projected local regression on the sphere. Local linear-type estimators have been recently used by Alonso-Pena et al. (2020) in order to propose no-effect and ANCOVA tests for regression models with circular response and/or covariate. However, the problem of assessing a certain class of parametric regression models with circular response and Euclidean covariates (up to the knowledge of the author), has not been considered in the statistical literature yet, neither for independent nor for spatially dependent observations.

In this chapter, new approaches for testing a linear-circular parametric regression model (circular response and $\mathbb{R}^d$-valued covariate) are proposed and empirically analyzed, both for independent and spatially correlated errors. Following similar ideas to those used in Chapter 2, the test statistics considered in these procedures are based on a comparison between a (non-smoothed or smoothed) parametric fit under the null

hypothesis and a nonparametric estimator of the circular regression function. More specifically, two different test statistics are designed. In the first one, a parametric estimator of the regression function under the null hypothesis is directly used, while in the second one, a smoothed version of this estimator is employed. Notice that, in this framework, a suitable measure of circular distance must be employed (see Jammalamadaka and SenGupta, 2001, Section 1.3.2). The null hypothesis that the regression function belongs to a certain parametric family is rejected if the distance between both fits exceeds a certain threshold. To perform the parametric estimation, procedures based on least squares or maximum likelihood are used (see Fisher and Lee, 1992; Lund, 1999; Presnell et al., 1998). For the nonparametric alternative, local polynomial-type estimators given in (3.14) are considered.

For the practical application of the proposals, the test statistics must be accompanied by a calibration procedure. In this case, such a procedure is not based on the asymptotic distribution, given that the convergence to the limit distribution under the null hypothesis will presumably be too slow. Instead, bootstrap methods are designed and their performance is analyzed and compared employing numerical experiments. For independent data, standard resampling procedures adapted to the context of regression models with a circular response and Euclidean covariates are used: a parametric circular residual bootstrap (PCB) and a nonparametric circular residual bootstrap (NPCB). The PCB approach consists in using the residuals obtained from the parametric fit in the bootstrap algorithm. If the circular regression function belongs to the parametric family considered in the null hypothesis, then the residuals will tend to be quite *similar* to the theoretical errors and, therefore, it is expected that the PCB method has a good performance. Following the proposal by González-Manteiga and Cao (1993), the NPCB method aims to increase the power of the test and, for this purpose, the residuals obtained from the nonparametric fit are the ones employed in the bootstrap procedure. The previous resampling procedures (PCB and NPCB) for independent data must be properly adapted for handling spatial correlation. Two specific procedures for test calibration which take the spatial correlation into account are also introduced: a parametric spatial circular residual bootstrap (PSCB) and a nonparametric spatial circular residual bootstrap (NPSCB). Similarly to the PCB but now for spatially correlated errors, the PSCB considers

the residuals obtained from the parametric fit under the null hypothesis. The relevant difference between PCB and PSCB is that, in order to mimic the dependence structure of the errors, a spatial circular process is fit to the residuals in PSCB. Samples coming from the fit process are employed in the bootstrap algorithm. The steps followed in NPSCB are similar at those employed in PSCB, but the residuals are obtained from the nonparametric regression estimator.

This chapter is organized as follows. Section 5.2 is devoted to present some ideas of goodness-of-fit tests for parametric circular regression models with independent data. Parametric circular regression estimators employed in the test statistics are presented in Section 5.2.1. Section 5.2.2 introduces the proposed test statistics. A description of the calibration algorithms considered is given in Section 5.2.3. Section 5.2.4 contains a simulation study for assessing the performance of the tests when using the PCB and NPCB resampling approaches to approximate the sampling distribution of the test statistics. Section 5.2.5 illustrates the testing proposals with the blue periwinkle and sand hopper orientation datasets introduced above. The extension of the testing procedures for spatially correlated data is presented in Section 5.3. Section 5.3.1 contains bootstrap approaches to calibrate the tests in this spatial framework. A simulation study for assessing the performance of the tests using PSCB and NPSCB methods is provided in Section 5.3.2.

5.2 Goodness-of-fit tests for parametric circular regression models with independent data

Considering the regression model (3.3), the goal of this chapter is to propose and analyze empirically different testing procedures to assess the suitability of a general class of parametric circular regression models. That is, solving a testing problem similar to that given in (1.34), but taking into account that, in this case, the parametric (circular) family is $\mathcal{M}^c_{\boldsymbol{\beta}} = \{m^c_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}$, where $m^c_{\boldsymbol{\beta}}$ is a certain parametric circular regression model with parameter vector $\boldsymbol{\beta}$. The specific testing problem is formulated as:

$$H_0 : m^c \in \mathcal{M}^c_{\boldsymbol{\beta}} \qquad \text{vs.} \qquad H_a : m^c \notin \mathcal{M}^c_{\boldsymbol{\beta}}. \tag{5.1}$$

As pointed out in Section 5.1, this testing problem is tackled by comparing a (non-smoothed or smoothed) parametric fit with a nonparametric estimator of the circular regression function m^c, measuring the circular distance between both fits and employing this distance as a test statistic. The parametric estimation methods considered in this proposal are described in the following section. For the nonparametric approach, the kernel-type circular regression estimators proposed in (3.14) are employed. These estimators are used to design two different test statistics. In the first one, the parametric estimator of the regression function is directly used, while in the second one, a smoothed version of this estimator is employed.

Notice that in Chapter 2 of this dissertation, in order to solve the testing problem (1.34), the (non-smoothed) parametric fit was not considered to define the test statistic, and only the smoothed parametric was used. As pointed out there, unlike the parametric fit, the smoothed version of the parametric estimator has the same expected value as the nonparametric one, under H_0, and thus the asymptotic distribution of the corresponding test statistic would be easier to derive. In this chapter, we will focus on analyzing empirically the performance of the test and, consequently, both test statistics are compared and studied.

5.2.1 Parametric circular regression estimation

As mentioned in Section 5.1, our proposal requires a parametric estimator of the circular regression function m^c, once a parametric family is set as the null hypothesis. Notice that, for instance, the procedures based on least squares, such as those ones presented in Sections 1.1.1 and 1.1.2 for Euclidean data, are not appropriate when the response variable is of circular nature. Minimizing the sum of squared differences between the observed and predicted values may lead to erroneous results, since the squared difference is not an appropriate measure on the circle.

A circular analog to least squares regression for models with a circular response and a set of Euclidean covariates was presented by Lund (1999). Specifically, assuming that regression model (3.3) holds and $m^c \in \mathcal{M}^c_{\boldsymbol{\beta}}$, a parametric estimator of $m^c_{\boldsymbol{\beta}}$ is constructed obtaining an estimator of $\boldsymbol{\beta}$, namely $\hat{\boldsymbol{\beta}}$, and computing $m^c_{\hat{\boldsymbol{\beta}}}$. A parameter estimate of $\boldsymbol{\beta}$ could be obtained by minimizing the sum of the circular

distances between the observed and predicted values as follows:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left\{1 - \cos\left[\Theta_i - m^c_{\boldsymbol{\beta}}(\mathbf{X}_i)\right]\right\}. \tag{5.2}$$

An equivalent parameter estimator can be obtained using a maximum-likelihood approach (Lund, 1999). If it is assumed that the response variable (conditionally on $\mathbf{X}$) follows a von Mises distribution with mean direction given by $m^c_{\boldsymbol{\beta}}$ and concentration parameter κ, the maximum likelihood estimator of $m^c_{\boldsymbol{\beta}}$ maximizes the following expression

$$\sum_{i=1}^{n} \cos\left[\Theta_i - m^c_{\boldsymbol{\beta}}(\mathbf{X}_i)\right]. \tag{5.3}$$

Notice that the circular least squares estimator given in (5.2) also maximizes the expression (5.3) and, therefore, assuming a von Mises distribution, the circular least squares estimator coincides with the maximum likelihood estimator (for further details, see Lund, 1999). Given the maximum likelihood estimator of $\boldsymbol{\beta}$, the maximum likelihood estimator of κ is given by the solution to

$$A(\hat{\kappa}) = \frac{1}{n} \sum_{i=1}^{n} \cos\left[\Theta_i - m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)\right],$$

where $A(\kappa) = I_1(\kappa)/I_0(\kappa)$, being I_0 and I_1 the modified Bessel functions of the first kind with order zero and one, respectively. As indicated in Lund (1999), numerical solutions to $A^{-1}(x)$ can be found in Best and Fisher (1981).

Assuming that the response variable follows a von Mises distribution and considering the general class of models for the circular regression function $\mathcal{M}^c_{\boldsymbol{\beta}} = \{\mu_0 + g(\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X}), \mu_0 \in [0, 2\pi), \boldsymbol{\beta}_1 \in \mathbb{R}^d\}$, where g is a link function mapping the real line onto the circle, an iterative reweighted least squares algorithm can be used to compute the maximum likelihood estimators of κ, μ_0 and $\boldsymbol{\beta}_1$ (see Fisher and Lee, 1992; Lund, 1999). The extension of these results to the case of a generic parametric family has not been explicitly considered.

Although the assumption that the response variable follows a von Mises distribution is quite common, other circular distributions can be used in this context. For

example, considering a projected normal distribution allows to define general regression models, such as the PMLM (Presnell et al., 1998; Scapini et al., 2002). This class of models deals with directional observations as projections onto the unit circle of unobserved response vectors in a multivariate linear model. Considering these type of regression models, the estimation of the parameters can be performed using maximum likelihood methods employing iterative procedures. For further details on the estimation approach in this case, we refer to Presnell et al. (1998).

5.2.2 The test statistics

In this section, two tests statistics to address the testing problem (5.1) (that is, to check if the circular regression function belongs to a general class of parametric models) are proposed. The first approach considers a weighted circular distance between the nonparametric and parametric fits:

$$T_{n,p}^{c,1} = \int_{\mathcal{D}} \{1 - \cos[\hat{m}_{\mathbf{H}}^{c}(\mathbf{x}; p) - m_{\hat{\boldsymbol{\beta}}}^{c}(\mathbf{x})]\} w(\mathbf{x}) d\mathbf{x}, \tag{5.4}$$

for $p = 0, 1$, where w is a weight function that helps in mitigating possible boundary effects. The estimators $\hat{m}_{\mathbf{H}}^{c}(\mathbf{x}; p)$, for $p = 0, 1$, are the Nadaraya–Watson- or the local linear-type estimators of the circular regression function m^c, given in (3.14). As for the parametric estimator $m_{\hat{\boldsymbol{\beta}}}^{c}$, the approaches described in Section 5.2.1 can be used to compute (5.4).

The second test statistic is similar to the first one, but considering a smoothed version of the parametric fit:

$$T_{n,p}^{c,2} = \int_{\mathcal{D}} \{1 - \cos[\hat{m}_{\mathbf{H}}^{c}(\mathbf{x}; p) - \hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}^{c}(\mathbf{x}; p)]\} w(\mathbf{x}) d\mathbf{x}, \tag{5.5}$$

where $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}^{c}(\mathbf{x}; p)$, for $p = 0, 1$, are smoothed versions of the parametric estimator $m_{\hat{\boldsymbol{\beta}}}^{c}$, which are given by:

$$\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}^{c}(\mathbf{x}; p) = \operatorname{atan2}[\hat{m}_{1,\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p), \hat{m}_{2,\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)], \tag{5.6}$$

being

$$\hat{m}_{j,\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};0) = \begin{cases} \dfrac{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \sin[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)]}{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 1, \\ \dfrac{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x}) \cos[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)]}{\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i - \mathbf{x})} & \text{if } j = 2, \end{cases}$$

and

$$\hat{m}_{j,\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};1) = \begin{cases} \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \hat{\mathbf{S}} & \text{if } j = 1, \\ \mathbf{e}_1^{\mathrm{T}} (\boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \boldsymbol{\mathcal{X}}_{\mathbf{x}})^{-1} \boldsymbol{\mathcal{X}}_{\mathbf{x}}^{\mathrm{T}} \boldsymbol{\mathcal{W}}_{\mathbf{x}} \hat{\mathbf{C}} & \text{if } j = 2, \end{cases}$$

with $\hat{\mathbf{S}} = \{\sin[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_1)], \ldots, \sin[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_n)]\}^{\mathrm{T}}$ and $\hat{\mathbf{C}} = \{\cos[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_1)], \ldots, \cos[m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_n)]\}^{\mathrm{T}}$. It should be noted that although the methodological developments are presented for the cases of $p = 0$ (Nadaraya–Watson) and $p = 1$ (local linear), they can be generalized to a higher polynomial degree p.

In order to formally address problem (5.1) using the test statistics $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$ given in (5.4) and in (5.5), respectively, it is essential to approximate the distribution of these test statistics under the null hypothesis. Deriving the asymptotic distribution of the statistics is out of the scope of this dissertation. However, some guidelines to compute these expressions are provided in Section 6.4. For the application in practice of our proposal, the distribution of the tests under the null hypothesis is approximated using bootstrap procedures and analyzed through an empirical study. Based on the practical results of Chapters 3 and 4, where the local linear-type estimator $m^c_{\mathbf{H}}(\mathbf{x};1)$ presented in general a slightly better performance than the Nadaraya–Watson one, just results corresponding to $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ are provided in this chapter.

Notice that if the null hypothesis in the testing problem given in (5.1) holds, then the (non-smoothed or smoothed) parametric fit and the nonparametric circular regression estimator will be similar and, therefore, the value of the test statistics $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$ will be small. Conversely, if the null hypothesis does not hold, the fits will be different and the value of $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$ will be large. So, the null hypothesis will be rejected if the circular distance between the parametric and the nonparametric fits exceeds a critical value.

For a visual illustration of the performance of the tests (for simplicity, a model

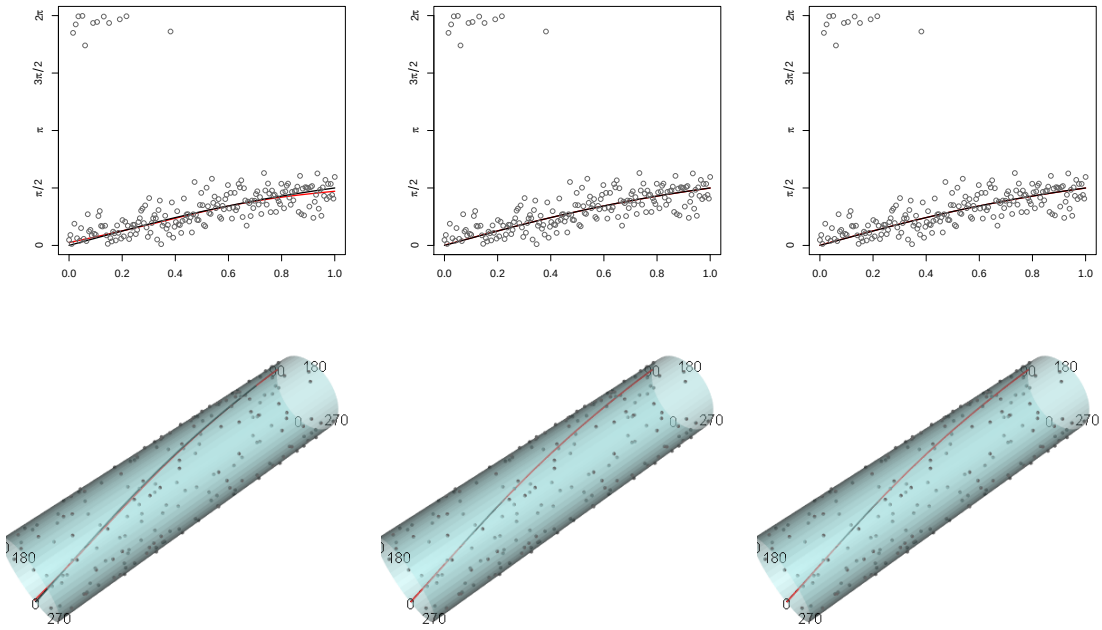

Figure 5.1: Linear (top panels) and cylinder (bottom panels) representations. Red lines: local linear-type regression estimator (left panels), parametric fit (center panels) and smoothed version of the parametric fit (right panels), with sample points and circular regression function (black lines). Equally-spaced sample of size $n = 200$ generated on the unit interval $\mathcal{D} = [0, 1]$, following model (3.3), with regression function (5.8), for $c = 0$, and circular errors ε_i drawn from a $vM(0, 10)$.

with a single covariate, that is, $d = 1$, is initially employed), consider an equally-spaced sample of size $n = 200$, generated in the unit interval $\mathcal{D} = [0, 1]$, following model (3.3), with regression function (5.8) and $c = 0$. The random errors ε_i are drawn from a von Mises distribution $vM(0, 10)$, with density function given in (1.40). If we want to test if $m^c(X) \in \mathcal{M}^c_{1,\boldsymbol{\beta}} = \{\mu_0 + 2\text{atan}(\beta_1 X), \mu_0 \in [0, 2\pi), \beta_1 \in \mathbb{R}\}$, using the test statistics given in (5.4) and in (5.5), for $p = 1$, the local linear-type regression estimator, $\hat{m}^c_h(x; 1)$, given in (3.14), as well as a parametric fit, $m^c_{\hat{\boldsymbol{\beta}}}(x)$, and its smoothed version $\hat{m}^c_{h,\hat{\boldsymbol{\beta}}}(x; 1)$ (denoting by h the bandwidth parameter when $d = 1$) must be computed. In this case, the estimator obtained from (5.3) is considered for the parametric fit. A triweight kernel and the optimal bandwidth obtained by minimizing the CASE, given in (3.34), are considered to compute $\hat{m}^c_h(x; 1)$ and

$\hat{m}^c_{h,\hat{\boldsymbol{\beta}}}(x;1)$.

Figure 5.1 shows the linear (top panels) and cylinder (bottom panels) representations of the estimates. In red lines the local linear-type regression estimator (left panel), the parametric fit (center panel) and the smoothed version of the parametric fit (right panel), with sample points and the circular regression function (black lines). All estimates show a very similar behaviour and, therefore, the value of the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ would presumably be small. Consequently, there may be no evidences against the assumption that the circular regression function belongs to the parametric family $\mathcal{M}^c_{1,\boldsymbol{\beta}}$.

A similar visual experiment considering a regression model with a circular response and two covariates is also presented. A sample of size $n = 400$ is generated on a regular grid in the unit square $\mathcal{D} = [0,1] \times [0,1]$, assuming the linear-circular regression model (3.3), with regression function (5.10) and $c = 0$. The random errors ε_i are also drawn from a von Mises distribution $vM(0,10)$. In this case, as in the previous example, in order to test if $m^c(\mathbf{X}) \in \mathcal{M}^c_{2,\boldsymbol{\beta}} = \{\mu_0 + 2\text{atan}(\beta_1 X_1 + \beta_2 X_2), \mu_0 \in [0, 2\pi), \beta_1, \beta_2 \in \mathbb{R}\}$, being $\mathbf{X} = (X_1, X_2)$, using the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$, the estimator obtained from (5.3) is employed for the parametric fit. The local linear-type estimator and the smoothed parametric fit are computed using a multiplicative triweight kernel and an optimal bandwidth obtained by minimizing the CASE, given in (3.34).

Figure 5.2 shows the theoretical circular regression function (top left panel), the local linear-type regression estimator (top right panel), the parametric fit (bottom left panel) and the smoothed version of the parametric fit (bottom right panel). It can be observed that estimates at top right, bottom left and bottom right panels seem to be very similar and, therefore, analogous conclusions to those given for $d = 1$, but in this case for $d = 2$, can be derived.

Notice that the test statistics given in (5.4) and in (5.5), as it happened in the test proposed in (2.1) for Euclidean data, require a $d \times d$ bandwidth matrix $\mathbf{H}$ (or a bandwidth parameter h, if $d = 1$). Similarly to Chapter 2, the performance of the proposed test statistics is analyzed for a range of bandwidths, in order to evaluate the impact of this parameter in the numerical results (both in the simulation study and in the real data illustrations).

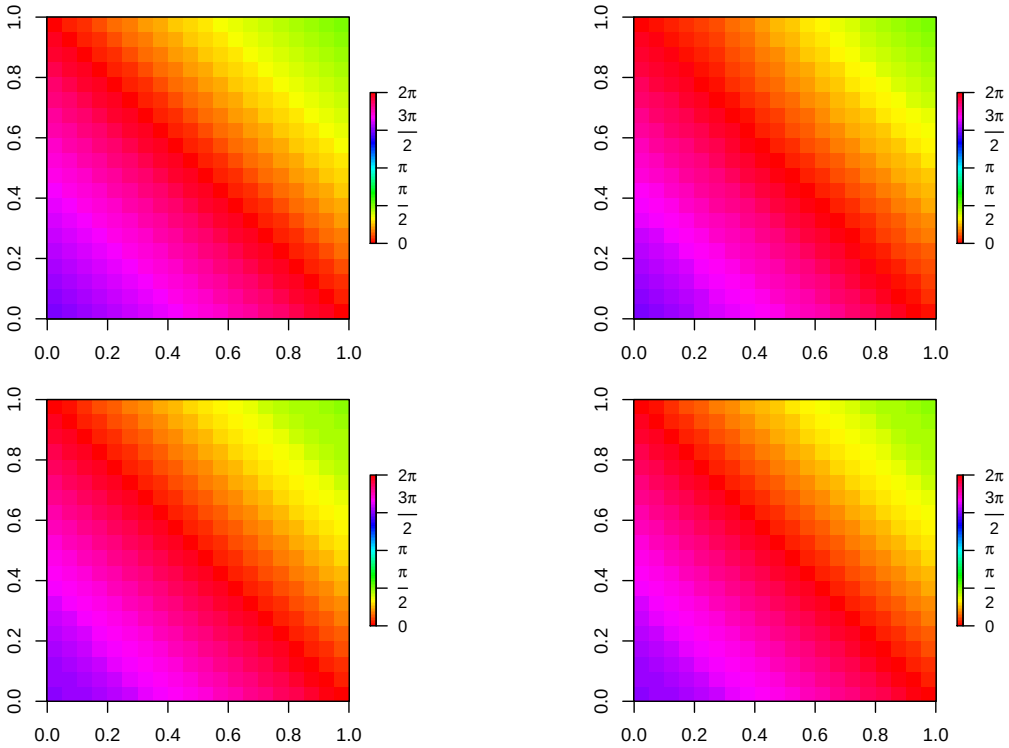

Figure 5.2: Circular regression function (top left panel), the local linear-type regression estimator (top right panel), the parametric fit (bottom left panel) and the smoothed version of the parametric fit (bottom right panel). Sample of size $n = 400$ generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, following model (3.3), with regression function (5.10), for $c = 0$, and circular errors ε_i drawn from a $vM(0, 10)$.

5.2.3 Calibration in practice

Once a suitable test statistic is available, in order to solve the testing problem (5.1), a procedure for calibration of critical values is required. This task can be done by means of bootstrap resampling algorithms.

In what follows, a description of two bootstrap proposals (PCB and NPCB) designed to approximate the distribution (under the null hypothesis) of the tests statistics, given in (5.4) and in (5.5), is presented. The main difference between them is the mechanism employed to obtain the residuals. As pointed out in Section 5.1, the residuals used in PCB come from the parametric regression estimator. On the other hand, for the NPCB algorithm, the residuals employed in the resampling process

are obtained from the nonparametric regression estimator. In order to present the PCB and NPCB resampling methods, a generic bootstrap algorithm is described. No matter the method used, $\hat{m}^c$ denotes the parametric or the nonparametric circular regression estimator.

Algorithm 2

1. Compute the nonparametric or the parametric regression estimates (described in Sections 3.3.1 and 5.2.1, respectively), namely $\hat{m}^c(\mathbf{X}_i)$, $i = 1, \ldots, n$, depending if a parametric (PCB) or a nonparametric (NPCB) bootstrap procedure is employed.
2. From the residuals $\hat{\varepsilon}_i = [\Theta_i - \hat{m}^c(\mathbf{X}_i)](\mathtt{mod}\, 2\pi)$, $i = 1, \ldots, n$, draw independent bootstrap residuals, $\hat{\varepsilon}_i^*$, $i = 1, \ldots, n$.
3. Obtain bootstrap samples $\{(\mathbf{X}_i, \Theta_i^*)\}_{i=1}^n$ with $\Theta_i^* = [m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) + \hat{\varepsilon}_i^*](\mathtt{mod}\, 2\pi)$, being $m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)$ the parametric regression estimator under H_0.
4. Using the bootstrap sample $\{(\mathbf{X}_i, \Theta_i^*)\}_{i=1}^n$, the bootstrap test statistics $T_{n,p}^{c,l,*}$, with $l = 1, 2$, are computed as in (5.4) and in (5.5).
5. Repeat Steps 2-4 a large number of times B.

The PCB is similar to the PB discussed in Section 2.3.1 for spatially correlated Euclidean data, but, in this case, for independent and circular observations. In Step 1 of the previous algorithm, the circular regression function is estimated parametrically, employing one of the procedures described in Section 5.2.1. Alternatively, the NPCB tries to avoid possible misspecification problems, using more flexible regression estimation methods than those employed in PCB. Then, following the same arguments as in González-Manteiga and Cao (1993) to increase the power of the test, in the NCPB method, the nonparametric circular regression estimator given in (3.14) is employed in Step 1 of the bootstrap Algorithm 2.

Notice that the empirical distribution of the B bootstrap test statistics can be employed to approximate the finite sample distribution of the test statistics $T_{n,p}^{c,1}$ and $T_{n,p}^{c,2}$ under the null hypothesis. Denoting by $\{T_{n,p,1}^{c,l,*}, \cdots, T_{n,p,B}^{c,l,*}\}$ (for $l = 1, 2$) the sample of the B bootstrap test statistics, given in (5.4) and in (5.5), and defining its $(1-\alpha)$ quantile $t_{\alpha,p}^{c,l,*}$, the null hypothesis in (5.1) will be rejected if $T_{n,p}^{c,l} > t_{\alpha,p}^{c,l,*}$. Additionally, as it was pointed out for Euclidean data in Chapter 2, the p-values of the test statistics can be approximated as in (2.7).

5.2.4 Simulation study

The finite sample performance of the proposed tests, using the bootstrap approaches described in Algorithm 2 for their calibration, is illustrated in what follows with a simulation study, considering a regression model with a single real-valued covariate and also with a bidimensional one. As pointed out previously, only results for $p = 1$ are presented in this section.

Simulation experiment with a single covariate

In order to study empirically the performance of the proposed tests considering a regression model with a circular response and a single real-valued covariate, the parametric regression family

$$\mathcal{M}^c_{1,\boldsymbol{\beta}} = \{\mu_0 + 2\text{atan}(\beta_1 X), \mu_0 \in [0, 2\pi), \beta_1 \in \mathbb{R}\} \tag{5.7}$$

is chosen, and for different values of c, the regression function

$$m^c_1(X) = 2\text{atan}(X) + c\,\text{asin}(2X^5 - 1) \tag{5.8}$$

is considered. Therefore, the parameter c controls whether the null ($c = 0$) or the alternative ($c \neq 0$) hypotheses hold in problem (5.1). Values $c = 0$, 1, and 2 are considered in the study. This circular regression function was plotted in Figure 5.1 (black lines) considering $c = 0$. For each value of c, 500 samples of sizes $n = 50, 100$ and 200 are generated on the unit interval $\mathcal{D} = [0, 1]$, considering an equally-spaced explanatory variable X, following model (3.3) with regression function (5.8). The independent circular errors ε_i are drawn from a von Mises distribution $vM(0, \kappa)$, with density function given in (1.40), for different values of the concentration ($\kappa = 5, 10$ and 15).

To analyze the behavior of the test statistics given in (5.4) and in (5.5), for $p = 1$, in the different scenarios, the bootstrap procedures described in Section 5.2.3 are applied, using $B = 500$ replications. The non-smoothed or smoothed parametric fits used for constructing (5.4) and (5.5) are computed using the estimators obtained from (5.3) and (5.6), respectively. The nonparametric fit is obtained using the estimator

given in (3.14), for $p = 1$, with a triweight kernel. We address the bandwidth selection problem by using the same procedure as the one used in Härdle and Mammen (1993), Alcalá et al. (1999), Opsomer and Francisco-Fernández (2010), among others, applying the tests on a grid of several bandwidths. In order to use a reasonable grid of bandwidths, the optimal bandwidth selected by minimizing the CASE given in (3.34), for $d = 1$, is calculated for each sample and for each scenario. In this case, the values of the average of the CASE optimal bandwidths are in the interval $[0.2, 0.6]$. Therefore, the values of the bandwidth parameter $h = 0.15, 0.25, 0.35, 0.45, 0.55, 0.65$, are considered to compute both test statistics (5.4) and (5.5). The weight function used in both tests is $w(x) = \mathbb{I}_{\{x\in[1/\sqrt{n},1-1/\sqrt{n}]\}}$, to avoid possible boundary effects.

Effect of sample size. Proportions of rejections of the null hypothesis, for a significance level $\alpha = 0.05$, considering $\kappa = 10$, and different sample sizes, are shown in Tables 5.1 and 5.2, when using $T_{n,1}^{c,1}$ and $T_{n,1}^{c,2}$, respectively. If $c = 0$ (null hypothesis), the proportions of rejections are similar to the theoretical level, although they are quite affected by the value of h. The tests attain the nominal significance level of 5%, since for appropriate values of h, the majority of proportions of rejections under the null hypothesis lie within the intervals $(0, 0.110)$, $(0.007, 0.093)$ and $(0.020, 0.080)$, when $n = 50, 100$ and 200, respectively. For alternative assumptions ($c = 1$ and $c = 2$), as expected, as the sample size increases the proportions of rejections are larger. For all the scenarios, the power of the tests becomes larger as the value of c increases. Notice that, in most of the cases, an increasing power of the tests when the values of h decrease is observed. It should be noted that NPCB presents a slightly better performance than the PCB. On the other hand, although both test statistics provide a similar behavior of the testing procedure, $T_{n,1}^{c,2}$ seems to give slightly better results.

Effect of κ. The performance of the tests $T_{n,1}^{c,1}$ and $T_{n,1}^{c,2}$ (for $\alpha = 0.05$) is studied for $n = 200$ and for different values of the concentration parameter κ. Results are included in Tables 5.3 and 5.4, when using $T_{n,1}^{c,1}$ and $T_{n,1}^{c,2}$, respectively. If $c = 0$, the proportions of rejections are similar to the theoretical level when using both bootstrap approaches (PCB and NPCB). For alternative assumptions, as expected, large values of the concentration parameter κ lead to an increase in power, which justifies the correct performance of the bootstrap procedures.

c	n	Method	$h = 0.15$	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$
0	50	PCB	0.032	0.036	0.034	0.040	0.042	0.042
		NPCB	0.048	0.040	0.044	0.048	0.046	0.050
	100	PCB	0.026	0.026	0.032	0.032	0.036	0.034
		NPCB	0.028	0.028	0.032	0.034	0.036	0.034
	200	PCB	0.024	0.028	0.028	0.034	0.036	0.034
		NPCB	0.026	0.034	0.026	0.036	0.040	0.046
1	50	PCB	0.100	0.124	0.148	0.162	0.156	0.152
		NPCB	0.142	0.156	0.170	0.184	0.184	0.174
	100	PCB	0.212	0.264	0.300	0.324	0.318	0.304
		NPCB	0.250	0.306	0.344	0.352	0.352	0.336
	200	PCB	0.504	0.604	0.642	0.660	0.668	0.666
		NPCB	0.548	0.636	0.674	0.686	0.692	0.680
2	50	PCB	0.380	0.506	0.574	0.606	0.618	0.598
		NPCB	0.478	0.582	0.638	0.672	0.678	0.670
	100	PCB	0.856	0.934	0.952	0.958	0.962	0.962
		NPCB	0.896	0.944	0.964	0.972	0.970	0.970
	200	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.1: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{1,\beta}$ with different sample sizes and $\kappa = 10$. The test statistic $T^{c,1}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	n	Method	$h = 0.15$	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$
0	50	PCB	0.032	0.032	0.040	0.044	0.042	0.040
		NPCB	0.044	0.042	0.046	0.050	0.052	0.050
	100	PCB	0.026	0.028	0.038	0.032	0.030	0.030
		NPCB	0.026	0.030	0.038	0.036	0.036	0.036
	200	PCB	0.028	0.028	0.028	0.024	0.032	0.038
		NPCB	0.026	0.030	0.030	0.022	0.032	0.038
1	50	PCB	0.106	0.118	0.144	0.140	0.156	0.146
		NPCB	0.140	0.154	0.168	0.182	0.188	0.180
	100	PCB	0.214	0.260	0.288	0.290	0.290	0.270
		NPCB	0.248	0.298	0.324	0.334	0.336	0.312
	200	PCB	0.502	0.582	0.610	0.618	0.644	0.620
		NPCB	0.536	0.610	0.632	0.650	0.654	0.640
2	50	PCB	0.380	0.500	0.548	0.570	0.562	0.558
		NPCB	0.476	0.574	0.620	0.626	0.638	0.620
	100	PCB	0.840	0.924	0.944	0.946	0.948	0.944
		NPCB	0.894	0.944	0.962	0.966	0.960	0.958
	200	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	0.998

Table 5.2: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{1,\beta}$ with different sample sizes and $\kappa = 10$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	κ	Method	$h = 0.15$	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$
0	5	PCB	0.030	0.030	0.024	0.024	0.028	0.030
		NPCB	0.034	0.030	0.024	0.024	0.028	0.030
	10	PCB	0.024	0.028	0.028	0.034	0.036	0.034
		NPCB	0.026	0.034	0.026	0.036	0.040	0.046
	15	PCB	0.026	0.034	0.030	0.034	0.038	0.040
		NPCB	0.026	0.030	0.028	0.034	0.034	0.036
1	5	PCB	0.200	0.262	0.282	0.306	0.290	0.278
		NPCB	0.216	0.276	0.306	0.314	0.320	0.294
	10	PCB	0.504	0.604	0.642	0.660	0.668	0.666
		NPCB	0.548	0.636	0.674	0.686	0.692	0.680
	15	PCB	0.764	0.836	0.878	0.880	0.868	0.850
		NPCB	0.784	0.856	0.882	0.882	0.872	0.868
2	5	PCB	0.872	0.916	0.930	0.942	0.942	0.928
		NPCB	0.884	0.918	0.938	0.946	0.940	0.930
	10	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000
	15	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.3: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{1,\beta}$ with different values of κ and $n = 200$. The test statistic $T^{c,1}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	κ	Method	$h = 0.15$	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$
0	5	PCB	0.032	0.028	0.022	0.022	0.028	0.026
		NPCB	0.034	0.028	0.026	0.030	0.030	0.028
	10	PCB	0.028	0.028	0.028	0.024	0.032	0.038
		NPCB	0.026	0.030	0.030	0.022	0.032	0.038
	15	PCB	0.028	0.034	0.038	0.034	0.034	0.034
		NPCB	0.026	0.038	0.036	0.036	0.032	0.034
1	5	PCB	0.198	0.252	0.264	0.272	0.268	0.250
		NPCB	0.218	0.274	0.280	0.290	0.280	0.264
	10	PCB	0.502	0.582	0.610	0.618	0.644	0.620
		NPCB	0.536	0.610	0.632	0.650	0.654	0.640
	15	PCB	0.752	0.826	0.862	0.868	0.868	0.858
		NPCB	0.782	0.846	0.874	0.874	0.884	0.868
2	5	PCB	0.870	0.910	0.918	0.932	0.928	0.918
		NPCB	0.884	0.916	0.930	0.942	0.938	0.932
	10	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	0.998
	15	PCB	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.4: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{1,\beta}$ with different values of κ and $n = 200$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

Simulations experiment with several covariates

The extension for regression models with a circular response and two covariates is analyzed in this section. For this purpose, the parametric regression family

$$\mathcal{M}^c_{2,\boldsymbol{\beta}} = \{\mu_0 + 2\text{atan}(\beta_1 X_1 + \beta_2 X_2), \mu_0 \in [0, 2\pi), \beta_1, \beta_2 \in \mathbb{R}\} \tag{5.9}$$

is chosen, and for different values of c the regression function

$$m^c_2(\mathbf{X}) = 2\text{atan}(-X_1 + X_2) + c\,\text{asin}(2X_1^3 - 1), \tag{5.10}$$

being $\mathbf{X} = (X_1, X_2)$, is considered. This circular regression function was plotted in Figure 5.2 (top left panel) considering $c = 0$. For each value of c ($c = 0$, 1, and 2), 500 samples of sizes $n = 100, 225$ and 400 are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, following model (3.3), with regression function (5.10). The circular errors ε_i are drawn from a von Mises distribution $vM(0, \kappa)$, with density function given in (1.40), for $\kappa = 5, 10$ and 15. The bootstrap procedures described in Section 5.2.3 are applied, using $B = 500$ replications. The non-smoothed or smoothed parametric fits used for constructing (5.4) and (5.5) are computed using the estimators obtained from (5.3) and (5.6), respectively. The nonparametric fit is obtained using the estimator given in (3.14), for $p = 1$, with a multiplicative triweight kernel. In order to simplify the calculations, the bandwidth matrix is restricted to a class of diagonal matrices with both equal elements. In this case, the diagonal elements of the average of the CASE optimal bandwidths are in the interval $[0.3, 0.8]$. Therefore, diagonal bandwidth matrices $\mathbf{H} = \text{diag}(h, h)$ with different values of h, $h = 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85$, are considered to compute both test statistics (5.4) and (5.5). The weight function used in both tests is $w(\mathbf{x}) = \mathbb{I}_{\{\mathbf{x}\in[1/\sqrt{n},1-1/\sqrt{n}]\times[1/\sqrt{n},1-1/\sqrt{n}]\}}$.

Effect of sample size. Proportions of rejections of the null hypothesis, for a significance level $\alpha = 0.05$, considering $\kappa = 10$, and different sample sizes, are shown Tables 5.5, when using $T^{c,1}_{n,1}$. It can be observed that using both bootstrap methods (PCB and NPCB), the tests have a reasonable behavior. If $c = 0$ (null hypothesis), the tests preserve the nominal significance level, since most of the proportions of rejections lie within the intervals $(0.007, 0.093)$, $(0.022, 0.078)$ and

$(0.029, 0.071)$, when $n = 100, 225$ and 400, respectively. For alternative assumptions ($c = 1$ and $c = 2$), NPCB presents a slightly better performance than the PCB. Notice that, in most of the cases, an increasing power of the tests when the values of h increase is observed. For all the scenarios, the power of the tests becomes larger as the value of c increases. Analogous conclusions at those given for $T_{n,1}^{c,1}$ were obtained when the test statistic $T_{n,1}^{c,2}$ was employed (Table 5.6).

Effect of κ. The performance of the bootstrap procedures is analyzed for $n = 400$ and for different values of the concentration parameter κ when using $T_{n,1}^{c,1}$, for $\alpha = 0.05$, in Table 5.7. If $c = 0$, the proportions of rejections are similar to the theoretical level when using both bootstrap approaches (PCB and NPCB). It can be observed that for larger values of the concentration parameter κ, the bandwidth values providing an effective calibration must be smaller. For alternative assumptions, if the value of the concentration parameter κ is larger, an increasing power is obtained. Results considering the test statistic $T_{n,1}^{c,2}$ are summarized in Table 5.8. Similar conclusions to those provided for $T_{n,1}^{c,1}$ were obtained.

c	n	Method	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$	$h = 0.75$	$h = 0.85$
0	100	PCB	0.030	0.034	0.048	0.050	0.062	0.066	0.068
		NPCB	0.044	0.048	0.058	0.062	0.062	0.064	0.068
	225	PCB	0.030	0.032	0.028	0.038	0.042	0.042	0.042
		NPCB	0.024	0.030	0.032	0.038	0.042	0.044	0.044
	400	PCB	0.042	0.040	0.042	0.038	0.030	0.036	0.038
		NPCB	0.034	0.038	0.040	0.030	0.032	0.036	0.036
1	100	PCB	0.102	0.066	0.034	0.008	0.004	0.004	0.000
		NPCB	0.158	0.106	0.038	0.014	0.004	0.004	0.000
	225	PCB	0.362	0.264	0.140	0.058	0.020	0.008	0.004
		NPCB	0.372	0.280	0.152	0.068	0.022	0.008	0.004
	400	PCB	0.724	0.614	0.396	0.198	0.066	0.030	0.020
		NPCB	0.722	0.616	0.392	0.190	0.076	0.032	0.020
2	100	PCB	0.574	0.548	0.442	0.302	0.176	0.098	0.070
		NPCB	0.640	0.600	0.478	0.342	0.202	0.114	0.078
	225	PCB	0.992	0.990	0.976	0.916	0.776	0.638	0.472
		NPCB	0.992	0.994	0.980	0.924	0.804	0.664	0.508
	400	PCB	1.000	1.000	1.000	0.998	0.992	0.976	0.932
		NPCB	1.000	1.000	1.000	0.998	0.996	0.984	0.940

Table 5.5: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}^{c}$ with different sample sizes and $\kappa = 10$. The test statistic $T_{n,1}^{c,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	n	Method	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$	$h = 0.75$	$h = 0.85$
0	100	PCB	0.040	0.048	0.060	0.054	0.062	0.050	0.050
		NPCB	0.060	0.064	0.066	0.064	0.066	0.070	0.068
	225	PCB	0.032	0.032	0.030	0.038	0.046	0.044	0.046
		NPCB	0.038	0.034	0.032	0.044	0.048	0.044	0.042
	400	PCB	0.030	0.034	0.040	0.042	0.036	0.038	0.030
		NPCB	0.032	0.032	0.040	0.036	0.030	0.034	0.034
1	100	PCB	0.132	0.220	0.292	0.332	0.336	0.344	0.342
		NPCB	0.194	0.266	0.332	0.368	0.382	0.386	0.370
	225	PCB	0.398	0.554	0.636	0.672	0.680	0.682	0.672
		NPCB	0.418	0.552	0.640	0.670	0.678	0.676	0.662
	400	PCB	0.944	0.984	0.994	0.994	0.990	0.990	0.988
		NPCB	0.938	0.978	0.994	0.994	0.992	0.988	0.988
2	100	PCB	0.508	0.736	0.856	0.894	0.904	0.904	0.898
		NPCB	0.556	0.752	0.854	0.898	0.902	0.902	0.902
	225	PCB	0.980	0.996	0.998	1.000	1.000	1.000	1.000
		NPCB	0.980	0.996	0.998	0.998	1.000	1.000	1.000
	400	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.6: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different sample sizes and $\kappa = 10$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	κ	Method	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$	$h = 0.75$	$h = 0.85$
0	5	PCB	0.026	0.034	0.040	0.038	0.042	0.038	0.038
		NPCB	0.026	0.036	0.038	0.038	0.042	0.038	0.038
	10	PCB	0.042	0.040	0.042	0.038	0.030	0.036	0.038
		NPCB	0.034	0.038	0.040	0.030	0.032	0.036	0.036
	15	PCB	0.038	0.044	0.040	0.038	0.032	0.038	0.038
		NPCB	0.030	0.036	0.038	0.036	0.034	0.040	0.038
1	5	PCB	0.354	0.268	0.176	0.090	0.046	0.020	0.014
		NPCB	0.360	0.290	0.178	0.094	0.048	0.020	0.012
	10	PCB	0.724	0.614	0.396	0.198	0.066	0.030	0.020
		NPCB	0.722	0.616	0.392	0.190	0.076	0.032	0.020
	15	PCB	0.936	0.802	0.560	0.294	0.140	0.050	0.022
		NPCB	0.922	0.792	0.554	0.302	0.136	0.050	0.026
2	5	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.7: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different values of κ and $n = 400$. The test statistic $T^{c,1}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

c	κ	Method	$h = 0.25$	$h = 0.35$	$h = 0.45$	$h = 0.55$	$h = 0.65$	$h = 0.75$	$h = 0.85$
0	5	PCB	0.030	0.024	0.024	0.026	0.028	0.030	0.026
		NPCB	0.046	0.046	0.050	0.050	0.050	0.054	0.052
	10	PCB	0.030	0.034	0.040	0.042	0.036	0.038	0.030
		NPCB	0.032	0.032	0.040	0.036	0.030	0.034	0.034
	15	PCB	0.034	0.042	0.040	0.044	0.038	0.042	0.040
		NPCB	0.024	0.024	0.038	0.044	0.042	0.036	0.036
1	5	PCB	0.566	0.684	0.744	0.776	0.786	0.794	0.786
		NPCB	0.574	0.692	0.752	0.776	0.794	0.790	0.780
	10	PCB	0.944	0.984	0.994	0.994	0.990	0.990	0.988
		NPCB	0.938	0.978	0.994	0.994	0.992	0.988	0.988
	15	PCB	0.990	0.998	0.998	0.998	0.998	0.998	0.998
		NPCB	0.990	0.998	0.998	0.998	0.998	0.998	0.998
2	5	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	10	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	15	PCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		NPCB	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 5.8: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different values of κ and $n = 400$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 2 is used. Significance level: $\alpha = 0.05$.

5.2.5 Real data illustration

In order to illustrate the performance in practice of the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$, given in (5.4) and in (5.5), respectively, two datasets are used. Considering regression model (3.3) with a single real-valued covariate, the testing procedure is applied to the blue periwinkles dataset. For a bidimensional real-valued covariate, the sand hoppers dataset, previously introduced in Chapter 3, is employed to illustrate the proposed methodology.

Based on the simulation study, where both $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ presented a very similar behavior, only the test statistic $T^{c,2}_{n,1}$ was employed in these illustrations. Moreover, taking into account that NPCB presented a slightly better performance than the PCB in the simulations, only the NPCB resampling approach was used to calibrate the test.

Blue periwinkle data

The blue periwinkle dataset, mentioned in Section 5.1, is described in this section in more detail, to illustrate the application of the proposed goodness-of-fit test $T_{n,1}^{c,2}$ for a regression model with a single real-valued covariate. These data can be found in Table 1 of Fisher and Lee (1992), and are available in the `R` package `circular` (Lund et al., 2020).

Directions and distances moved by small blue periwinkles, after they had been transplanted downshore from the height at which they normally live, are considered. Figure 5.3 (left panel) shows the observations of this dataset, which were analyzed and modeled by different authors in the literature. In order to study how orientation varies with distance, in Fisher and Lee (1992), a parametric regression model was fit, considering that the regression function belongs to the parametric family $\mathcal{M}_{1,\boldsymbol{\beta}}^{c}$, given in (5.7). An iterative reweighted least squares algorithm to perform the maximum likelihood estimation of the parameters was employed. On the other hand, in Presnell et al. (1998), a parametric approach was also used to model these data. However, in this case, a PMLM was assumed, considering linear models on the covariate (distance) for the means of the bivariate normal distribution that is projected. Notice that if a projected normal distribution is assumed, with identity covariance matrix in (1.41), it holds that $\tan(\mu) = \mu_2/\mu_1$, being μ the circular mean direction, and μ_1 and μ_2 the mean components of the bivariate normal distribution that is projected (Presnell et al., 1998; Wang and Gelfand, 2013). Therefore, using this approach, the following parametric family is considered:

$$\mathcal{M}_{3,\boldsymbol{\beta}}^{c} = \{\text{atan2}(\beta_{0,2} + \beta_{1,2}\text{X}, \beta_{0,1} + \beta_{1,1}\text{X}), \quad \beta_{0,2}, \beta_{1,2}, \beta_{0,1}, \beta_{1,1} \in \mathbb{R}\},$$

where X represents the distance moved by the small blue periwinkles. This dataset was also explored by Di Marzio et al. (2013) using a nonparametric approach. Considering a regression model with a circular response (direction) and a single real-valued covariate (distance), the regression function was estimated using kernel-type methods.

In order to decide if $\mathcal{M}_{1,\boldsymbol{\beta}}^{c}$ or $\mathcal{M}_{3,\boldsymbol{\beta}}^{c}$ are plausible parametric models for the regression function with this dataset, the test statistic $T_{n,1}^{c,2}$ is applied twice considering

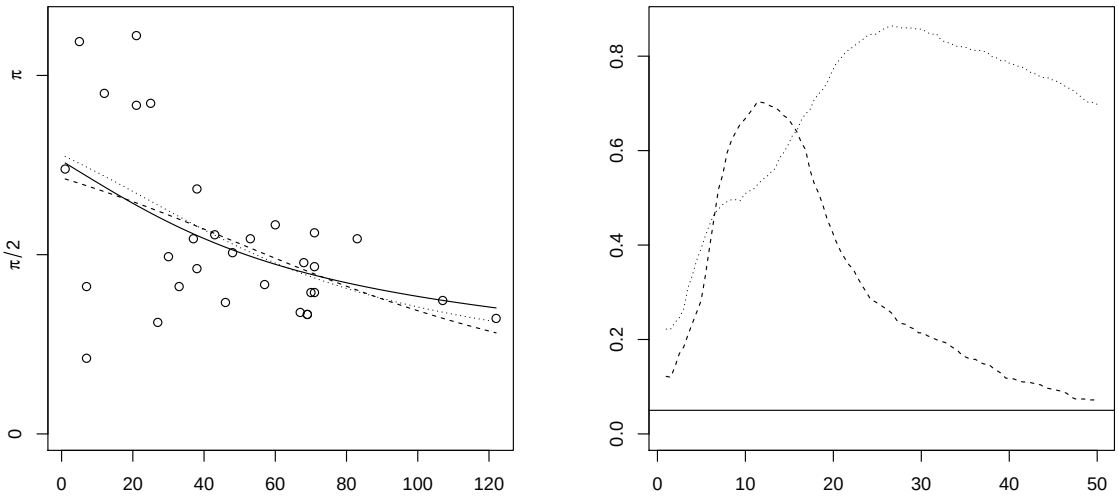

Figure 5.3: Left panel: Sample of directions and distances moved by periwinkles (circle points), smoothed versions of the parametric fits when considering the parametric families $\mathcal{M}^c_{1,\boldsymbol{\beta}}$ (dashed line) and $\mathcal{M}^c_{3,\boldsymbol{\beta}}$ (dotted line), and local linear-type regression estimator (solid line), using the CV^c bandwidth. Right panel: p-values of the test when considering the parametric family $\mathcal{M}^c_{1,\boldsymbol{\beta}}$ (dashed line) and $\mathcal{M}^c_{3,\boldsymbol{\beta}}$ (dotted line) as the null hypothesis for different values of h. Horizontal solid line represents the value 0.05.

$B = 500$ replications. In both cases, the parametric fits were computed by maximum likelihood (see Section 5.2.1). For further details on the estimation procedures, we refer to Fisher and Lee (1992) and Presnell et al. (1998). As for the nonparametric fit, the local linear-type estimator given in (3.14) with a triweight kernel was considered. As pointed out before, the performance of the test is analyzed in a range of bandwidths.

Figure 5.3 (left panel) shows the smoothed versions of the parametric fits when considering the parametric families $\mathcal{M}^c_{1,\boldsymbol{\beta}}$ (dashed line) and $\mathcal{M}^c_{3,\boldsymbol{\beta}}$ (dotted line), and the nonparametric regression estimator (solid line), using the CV^c bandwidth (see Di Marzio et al., 2013, for further details on bandwidth selection in this context). These curves are compared in the proposed test statistic. Figure 5.3 (right panel) shows the p-values of the tests when considering the parametric families $\mathcal{M}^c_{1,\boldsymbol{\beta}}$ (dashed line) or $\mathcal{M}^c_{3,\boldsymbol{\beta}}$ (dotted line) as the null hypothesis, using the significance trace (as it was performed in Chapter 2 for regression models with Euclidean response and

covariates, and spatially correlated errors). Taking into account this plot, there are no evidences to reject the null hypothesis in both testing problems. However, it can be observed that for h larger than 15, the p-value decreases considerably when considering the parametric family $\mathcal{M}^c_{1,\boldsymbol{\beta}}$ in the null hypothesis.

Sand hopper data

In this section, the testing procedure is applied to the sand hopper dataset, previously introduced in Chapter 3. With the purpose of analyzing how male and female sand hopper orientation behaves when other variables are included as covariates (such as azimuth, pressure, temperature, among others), both parametric and nonparametric approaches have been considered. For instance, following the proposal in Presnell et al. (1998), Scapini et al. (2002) employed a parametric approach, assuming a projected normal distribution for the scape directions. The authors considered linear models on the covariates for the means of the bivariate normal distribution that is projected. Using nonparametric tools, in Chapter 3 of this dissertation, this dataset (for males and females) was also explored in order to study how orientation behaves when temperature and (relative) humidity are included as covariates.

In order to determine if a parametric multiple regression model is an appropriate representation of these datasets (male and female sand hoppers), it is necessary to carry out a goodness-of-fit test for the selected parametric model. Similarly to Chapter 3, for illustration purposes, only observations corresponding to (relative) humidity values larger than 45% are considered. Assuming the parametric model used in Presnell et al. (1998) for these datasets, and taking into account the arguments in the previous section regarding the PMLM, the following parametric family is chosen:

$$\mathcal{M}^c_{4,\boldsymbol{\beta}} = \{\text{atan2}(\beta_{0,2} + \beta_{1,2}\text{X}_1 + \beta_{2,2}\text{X}_2, \beta_{0,1} + \beta_{1,1}\text{X}_1 + \beta_{2,1}\text{X}_2)\},$$

with $\beta_{0,2}, \beta_{1,2}, \beta_{2,2}, \beta_{0,1}, \beta_{1,1}, \beta_{2,1} \in \mathbb{R}$, and $X_1 =$ "temperature" and $X_2 =$ "humidity".

The test (5.5) is applied with $B = 500$ replications. The parametric fit was computed by maximum likelihood (for further details on the estimation procedure, we refer to Presnell et al., 1998). As for the nonparametric fit, the local linear-type estimator given in (3.14) with a multiplicative triweight kernel was considered. The

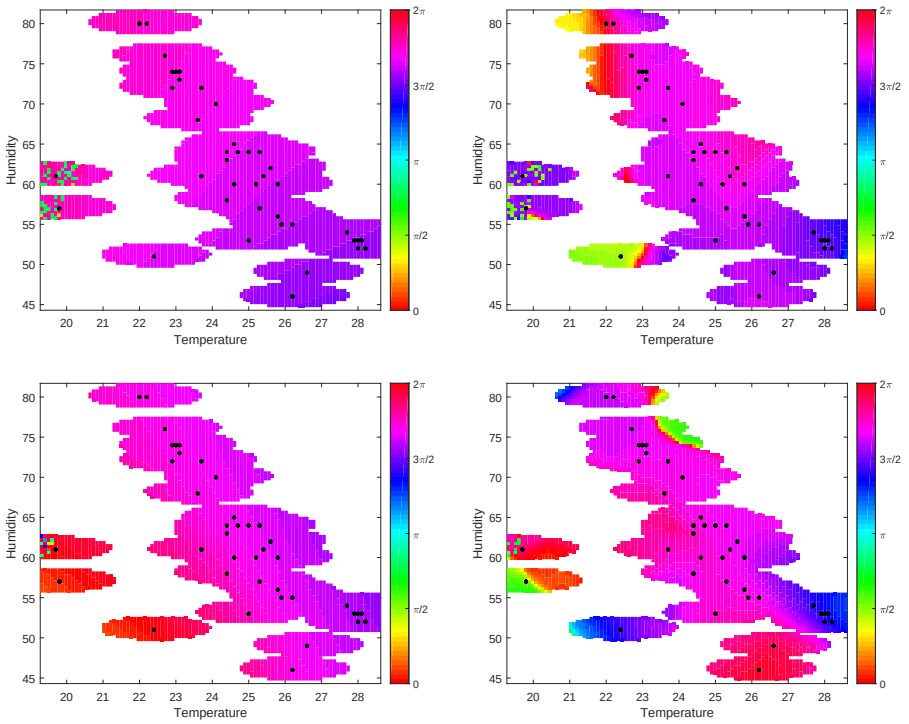

Figure 5.4: Smoothed version of the parametric fit for male (top left panel) and female (bottom left panel), and local linear-type regression estimators for male (top right panel) and for female (bottom right panel), using the CV^{c} bandwidth matrices given in (3.38) and in (3.39), for male and female, respectively. Horizontal axis: temperature in Celsius degrees. Vertical axis: relative humidity in percentage.

bandwidth was taken as a diagonal matrix $\mathbf{H} = \mathrm{diag}(h_1, h_2)$, being the values of h_1 and h_2 different. The range of bandwidths was selected taking into account the CV^{c} bandwidth matrices, given in (3.38) and in (3.39), for male and female, respectively.

Figure 5.4 shows the smoothed version of the parametric fit for male (top left panel) and female (bottom left panel), and the nonparametric regression estimators for male (top right panel) and for female (bottom right panel), using the CV^{c} bandwidth matrices given in (3.38) and in (3.39), for male and female, respectively. The plots corresponding to the left panels are compared with the right panels in the proposed test statistic. Figure 5.5 shows approximated the p-values of the test for male (left panel) and female (right panel), using the significance trace. Taking into account this figure, there are no evidences against the circular regression function

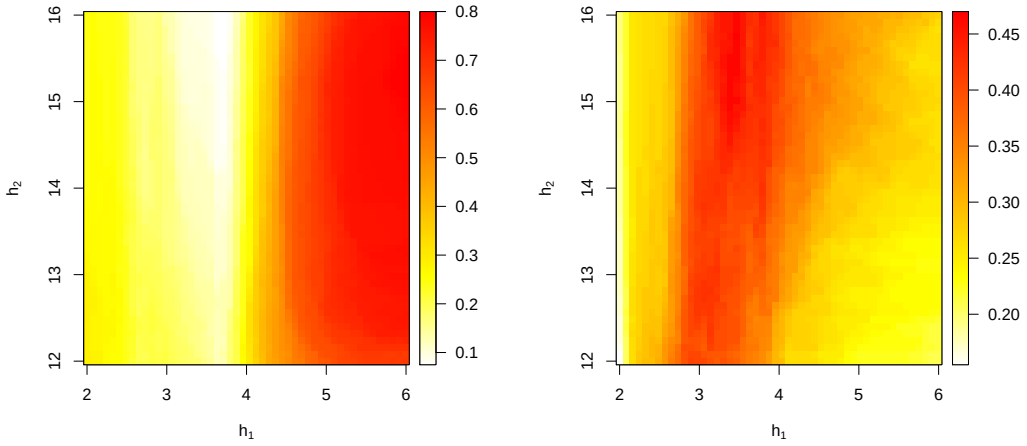

Figure 5.5: For male (left panel) and female (right panel) sand hopper orientation dataset, p-values of the test for different values of h_1 and h_2, considering the parametric family $\mathcal{M}^c_{4,\boldsymbol{\beta}}$ as the null hypothesis.

belonging to the parametric family $\mathcal{M}^c_{4,\boldsymbol{\beta}}$, for both sexes.

5.3 Goodness-of-fit tests for parametric circular regression models with spatially correlated data

The testing problem (5.1) is addressed in Section 5.2 for independent data, by constructing weighted test statistics. In this section, these test statistics are also analyzed considering a linear-circular regression model with spatially correlated errors. For this purpose, the estimators described in Section 5.2.1 are also employed for the parametric fit. Probably, more accurate results would be obtained if an estimator taking the spatial dependence structure into account was used. However, the problem of estimating parametrically the regression function accounting the dependence structure, up to the knowledge of the author, has not been tackled in the statistical literature. Some guidelines about a possible iterative least squares estimator (taking the possible spatial dependence structure into account) are provided in Section 6.4 of this dissertation. Kernel-type estimators given in (3.14) are employed for the nonparametric fit.

For illustration purposes, a sample of size $n = 400$ is generated on a bidimensional regular grid in the unit square, assuming the linear-circular regression model (3.3) with spatially correlated errors, and regression function (5.10), being $c = 0$. The errors are drawn from a wrapped Gaussian spatial process (Jona-Lasinio et al., 2012), described in Section 4.2.1, with zero mean and exponential covariance structure, given in (4.2), with $\sigma^2 = 1$ and $a_e = 0.3$. In order to test if the circular regression function can be assumed to belong to the parametric family $\mathcal{M}^c_{2,\boldsymbol{\beta}}$, given in (5.9), using the test statistics given in (5.4) and in (5.5), for $p = 1$, $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 1)$, $m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{x})$ and $\hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 1)$ fits must be computed. For the parametric fit, the estimator obtained from (5.3) is employed. A multiplicative triweight kernel and the optimal bandwidth obtained by minimizing the CASE, given in (3.34), of the local linear-type estimator are considered to compute $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; 1)$ and $\hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; 1)$. Figure 5.6 shows the theoretical circular regression function (top left panel), the local linear-type regression estimator (top right panel), the parametric fit (bottom left panel) and the smoothed version of the parametric fit (bottom right panel). It can be observed that estimates at top right, bottom left and bottom right panels seem to be very similar and, therefore, the value of the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ should be small. Consequently, the formal application of the tests will probably lead to assert that there is no evidences against the assumption that the regression function belongs to the parametric family $\mathcal{M}^c_{2,\boldsymbol{\beta}}$. Practical methods to calibrate the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$, given in (5.4) and in (5.5), for spatially correlated data are presented in the following section.

5.3.1 Calibration in practice

This section is devoted to present bootstrap resampling methods to calibrate in practice the test statistics $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$, given in (5.4) and in (5.5), respectively, considering the linear-circular regression model (3.3) with spatially correlated errors.

The bootstrap Algorithm 2, which was designed for independent data, should not be used for spatial processes, as it does not account for the correlation structure. The aim of this section is to describe two different proposals for test calibration which take the dependence of the data into account (PSCB and NPSCB). Similarly to the methods presented in Section 5.2.3, the main difference between them is how the resampling residuals (required for mimicking the dependence structure of the errors) are computed. In PSCB (similarly to PCB), the residuals are obtained

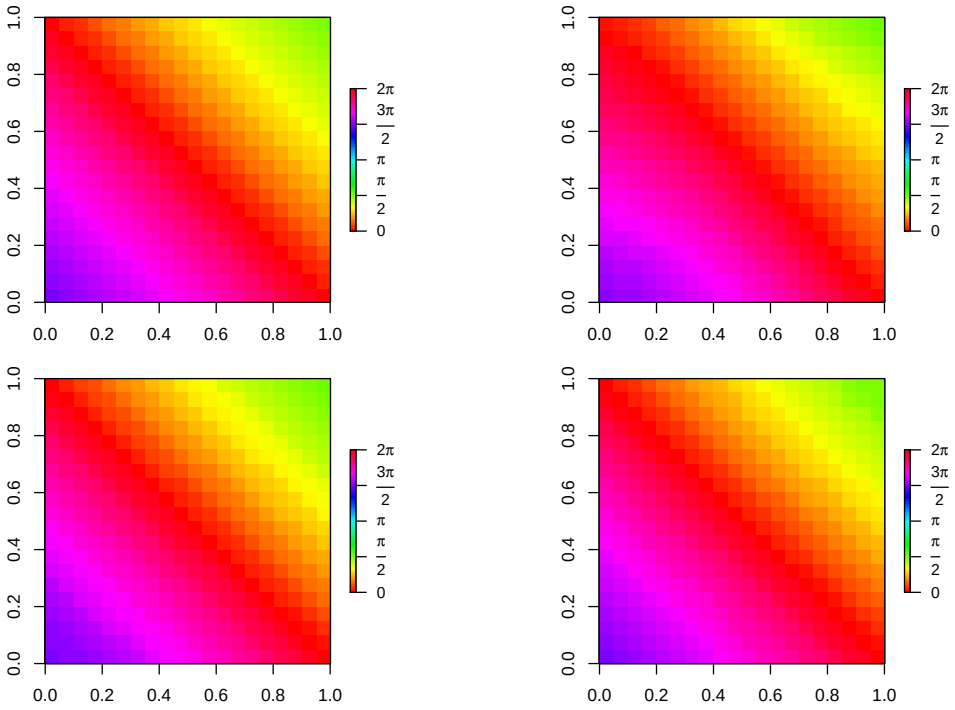

Figure 5.6: Circular regression function (top left panel), the local linear-type regression estimator (top right panel), the parametric fit (bottom left panel) and the smoothed version of the parametric fit (bottom right panel). Sample of size $n = 400$ generated on a regular grid in the unit square $\mathcal{D} = [0,1] \times [0,1]$, following model (3.3), with regression function (5.10), for $c = 0$, and circular errors ε_i drawn from from wrapped Gaussian spatial processes with zero mean and exponential covariance structure, given in (4.2), with $\sigma^2 = 1$ and $a_e = 0.3$.

from the parametric regression estimator, while in NPSCB (analogously to NPCB), the residuals are obtained from the nonparametric regression estimator. In both approaches, in order to mimic the dependence structure of the errors, an appropriate spatial circular process model is fit to the residuals.

Next, a generic bootstrap algorithm is introduced to present the PSCB and NPSCB resampling approaches. As in Algorithm 2, no matter the method used, either parametric or nonparametric, $\hat{m}^c$ denotes the parametric or the nonparametric circular regression estimator.

Algorithm 3

1. Compute the nonparametric or the parametric circular regression estimator (described in Sections 3.3.1 and 5.2.1, respectively), namely $\hat{m}^c(\mathbf{X}_i)$, $i = 1, \ldots, n$, depending if a parametric (PSCB) or a nonparametric (NPSCB) bootstrap procedure is employed.
2. From the residuals $\hat{\varepsilon}_i = [\Theta_i - \hat{m}^c(\mathbf{X}_i)](\mathtt{mod}\, 2\pi)$, $i = 1, \ldots, n$, fit a spatial circular process.
3. Generate a random sample from the fit model, $\hat{\varepsilon}_i^*$, $i = 1, \ldots, n$.
4. Obtain bootstrap samples $\{(\mathbf{X}_i, \Theta_i^*)\}_{i=1}^n$ with $\Theta_i^* = [m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i) + \hat{\varepsilon}_i^*](\mathtt{mod}\, 2\pi)$, being $m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{X}_i)$ the parametric regression estimator.
5. Using the bootstrap sample $\{(\mathbf{X}_i, \Theta_i^*)\}_{i=1}^n$, the bootstrap test statistics $T_{n,p}^{c,1,*}$ and $T_{n,p}^{c,2,*}$ are computed as in (5.4) and in (5.5), respectively.
6. Repeat Steps 3-5 a large number of times B.

Notice that Algorithm 3 is a modification of Algorithm 2. Two additional steps are included in Algorithm 3 (Steps 2 and 3) trying to mimic properly the distribution of spatial dependence structure of the circular errors in the bootstrap procedure.

As pointed out in Section 5.2.3 for independent data, considering the test statistics $T_{n,p}^{c,l}$ ($l = 1, 2$), given in (5.4) and in (5.5), the null hypothesis in (5.1) will be rejected if $T_{n,p}^{c,l} > t_{\alpha,p}^{c,l,*}$, where $t_{\alpha,p}^{c,l,*}$ is the $(1-\alpha)$ quantile of the sample of the B bootstrap test statistics $\{T_{n,p,1}^{c,l,*}, \cdots, T_{n,p,B}^{c,l,*}\}$. Moreover, the p-value of the test statistic can be approximated as in (2.7).

5.3.2 Simulation experiment

The performance of the proposed test statistics and the bootstrap procedures, described in Algorithm 3, are analyzed in a simulation study. As pointed out previously, only results for $p = 1$ are presented. The parametric circular regression family $\mathcal{M}^c_{2,\boldsymbol{\beta}}$, given in (5.9), is chosen, and for different values of c ($c = 0, 1, 2$), the regression function (5.10) is considered.

In this study, 500 samples of sizes $n = 100, 225$ and 400 are generated on a regular grid in the unit square $\mathcal{D} = [0, 1] \times [0, 1]$, assuming the linear-circular regression model (3.3), with regression function (5.10), but considering circular spatially correlated

errors generated from wrapped Gaussian spatial processes (Jona-Lasinio et al., 2012), described in Section 4.2.1. Specifically, the circular errors $\{\varepsilon_i,\ i = 1, \ldots, n\}$ are drawn considering the same scenarios as in Section 4.5, namely, realizations of zero mean processes with exponential covariance structure given in (4.2). The value of the variance σ^2 is fixed equal to one, and different values of the range parameter are considered: $a_e = 0.1, 0.3, 0.6$.

The performance of Algorithm 3 is analyzed in this section. Notice that Algorithm 2, which was designed for independent observations, should not be used in a spatial framework. In order to illustrate this issue (details of this algorithm were provided in Section 5.2.4), Tables 5.9 and 5.10 show the proportions of rejections of the null hypothesis for different sample sizes and $\alpha = 0.05$, when using $T_{n,1}^{c,1}$ and $T_{n,1}^{c,2}$, respectively, and the Algorithm 2 is applied. Considering both test statistics, it may seem that PCB and NPCB present a good behavior in terms of power, however, the proportions of rejections under the null hypothesis are very large. Results for $n = 400$ and different spatial dependence degrees (controlled by the range parameter, a_e) are summarized in Tables 5.11 and 5.12, when using $T_{n,1}^{c,1}$ and $T_{n,1}^{c,2}$, respectively. Again, it can be obtained that the tests do not work properly under the null hypothesis.

c	n	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	100	PCB	0.696	0.728	0.634	0.562	0.494	0.434	0.366	0.336
		NPCB	0.778	0.756	0.672	0.592	0.530	0.492	0.424	0.396
	225	PCB	0.996	0.984	0.966	0.928	0.882	0.832	0.772	0.722
		NPCB	0.996	0.988	0.966	0.932	0.894	0.852	0.790	0.742
	400	PCB	1.000	1.000	0.988	0.982	0.960	0.928	0.910	0.878
		NPCB	1.000	1.000	0.990	0.982	0.960	0.928	0.914	0.880
1	100	PCB	0.450	0.522	0.510	0.484	0.466	0.448	0.436	0.430
		NPCB	0.586	0.582	0.556	0.538	0.510	0.480	0.470	0.456
	225	PCB	0.742	0.666	0.616	0.566	0.548	0.534	0.520	0.504
		NPCB	0.798	0.732	0.662	0.622	0.594	0.578	0.554	0.550
	400	PCB	0.832	0.776	0.740	0.700	0.680	0.658	0.638	0.634
		NPCB	0.864	0.818	0.780	0.758	0.738	0.706	0.692	0.676
2	100	PCB	0.602	0.750	0.768	0.766	0.750	0.736	0.724	0.710
		NPCB	0.824	0.824	0.814	0.804	0.802	0.794	0.784	0.778
	225	PCB	0.746	0.752	0.740	0.726	0.722	0.706	0.680	0.678
		NPCB	0.802	0.798	0.790	0.780	0.772	0.760	0.734	0.728
	400	PCB	0.772	0.762	0.750	0.742	0.712	0.692	0.676	0.664
		NPCB	0.816	0.808	0.794	0.786	0.770	0.764	0.732	0.722

Table 5.9: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different sample sizes. Model parameters: $\sigma^2 = 1$ and $a_e = 0.3$. The test statistic $T_{n,1}^{c,1}$ is employed, and the Algorithm 2 for independent data is used. Significance level: $\alpha = 0.05$.

c	n	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	100	PCB	0.700	0.728	0.630	0.560	0.496	0.420	0.372	0.344
		NPCB	0.780	0.760	0.676	0.590	0.530	0.478	0.420	0.404
	225	PCB	0.996	0.986	0.968	0.926	0.876	0.844	0.788	0.742
		NPCB	0.996	0.986	0.968	0.932	0.890	0.850	0.806	0.774
	400	PCB	1.000	1.000	0.988	0.982	0.964	0.934	0.892	0.878
		NPCB	1.000	1.000	0.990	0.982	0.966	0.936	0.898	0.882
1	100	PCB	0.450	0.522	0.510	0.484	0.466	0.448	0.436	0.430
		NPCB	0.596	0.594	0.566	0.542	0.524	0.504	0.500	0.492
	225	PCB	0.746	0.684	0.638	0.602	0.586	0.568	0.568	0.558
		NPCB	0.808	0.740	0.684	0.642	0.616	0.610	0.600	0.606
	400	PCB	0.840	0.792	0.762	0.736	0.728	0.716	0.710	0.712
		NPCB	0.866	0.822	0.794	0.768	0.762	0.752	0.746	0.742
2	100	PCB	0.602	0.750	0.768	0.766	0.750	0.736	0.724	0.710
		NPCB	0.828	0.830	0.824	0.814	0.806	0.800	0.800	0.800
	225	PCB	0.752	0.762	0.756	0.742	0.738	0.730	0.724	0.722
		NPCB	0.806	0.804	0.796	0.790	0.788	0.780	0.776	0.774
	400	PCB	0.780	0.778	0.770	0.760	0.758	0.756	0.754	0.752
		NPCB	0.824	0.818	0.816	0.802	0.798	0.792	0.792	0.790

Table 5.10: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{12,\beta}$ with different sample sizes. Model parameters: $\sigma^2 = 1$ and $a_e = 0.3$. The test statistic $T_{n,1}^{c,2}$ is employed, and the Algorithm 2 for independent data is used. Significance level: $\alpha = 0.05$.

c	a_e	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	0.1	PCB	0.888	0.672	0.522	0.420	0.396	0.338	0.288	0.254
		NPCB	0.884	0.680	0.534	0.436	0.398	0.346	0.298	0.270
	0.3	PCB	1.000	1.000	0.988	0.982	0.960	0.928	0.910	0.878
		NPCB	1.000	1.000	0.990	0.982	0.960	0.928	0.914	0.880
	0.6	PCB	0.988	0.982	0.970	0.960	0.952	0.940	0.924	0.914
		NPCB	0.988	0.982	0.970	0.960	0.954	0.946	0.926	0.916
1	0.1	PCB	0.836	0.804	0.774	0.746	0.724	0.694	0.654	0.648
		NPCB	0.902	0.872	0.832	0.796	0.774	0.748	0.712	0.704
	0.3	PCB	0.832	0.776	0.740	0.700	0.680	0.658	0.638	0.634
		NPCB	0.864	0.818	0.780	0.758	0.738	0.706	0.692	0.676
	0.6	PCB	0.786	0.704	0.622	0.566	0.466	0.392	0.292	0.234
		NPCB	0.828	0.802	0.770	0.744	0.724	0.706	0.670	0.660
2	0.1	PCB	0.808	0.810	0.798	0.780	0.756	0.730	0.710	0.688
		NPCB	0.882	0.878	0.862	0.858	0.838	0.806	0.778	0.758
	0.3	PCB	0.772	0.762	0.750	0.742	0.712	0.692	0.676	0.664
		NPCB	0.816	0.808	0.794	0.786	0.770	0.764	0.732	0.722
	0.6	PCB	0.796	0.790	0.774	0.758	0.754	0.742	0.730	0.718
		NPCB	0.838	0.830	0.818	0.802	0.786	0.768	0.762	0.760

Table 5.11: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}^{c}$ with different values of a_e. Model parameters: $\sigma^2 = 1$ and $n = 400$. The test statistic $T_{n,1}^{c,1}$ is employed, and the Algorithm 2 for independent data is used. Significance level: $\alpha = 0.05$.

c	a_e	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	0.1	PCB	0.888	0.682	0.534	0.436	0.400	0.360	0.314	0.284
		NPCB	0.888	0.674	0.554	0.444	0.402	0.366	0.318	0.298
	0.3	PCB	1.000	1.000	0.988	0.982	0.964	0.934	0.892	0.878
		NPCB	1.000	1.000	0.990	0.982	0.966	0.936	0.898	0.882
	0.6	PCB	0.988	0.982	0.970	0.960	0.952	0.938	0.932	0.920
		NPCB	0.988	0.982	0.972	0.962	0.956	0.946	0.932	0.920
1	0.1	PCB	0.842	0.824	0.792	0.764	0.748	0.738	0.734	0.736
		NPCB	0.916	0.884	0.852	0.824	0.808	0.802	0.812	0.806
	0.3	PCB	0.840	0.792	0.762	0.736	0.728	0.716	0.710	0.712
		NPCB	0.866	0.822	0.794	0.768	0.762	0.752	0.746	0.742
	0.6	PCB	0.806	0.770	0.746	0.726	0.718	0.702	0.688	0.688
		NPCB	0.836	0.812	0.780	0.758	0.744	0.730	0.722	0.718
2	0.1	PCB	0.800	0.798	0.792	0.766	0.734	0.720	0.682	0.674
		NPCB	0.772	0.698	0.604	0.490	0.372	0.280	0.240	0.222
	0.3	PCB	0.780	0.778	0.770	0.760	0.758	0.756	0.754	0.752
		NPCB	0.824	0.818	0.816	0.802	0.798	0.792	0.792	0.790
	0.6	PCB	0.804	0.800	0.788	0.782	0.776	0.766	0.762	0.760
		NPCB	0.526	0.542	0.534	0.524	0.508	0.492	0.466	0.450

Table 5.12: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different values of a_e. Model parameters: $\sigma^2 = 1$ and $n = 400$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 2 for independent data is used. Significance level: $\alpha = 0.05$.

The bootstrap procedures described in Algorithm 3 are now applied using $B = 500$ replications. Again, the test statistics $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ given in (5.4) and in (5.5), are computed using the non-smoothed or smoothed parametric fits, obtained from (5.3) and (5.6), respectively, while the nonparametric fit is obtained employing the estimator given in (3.14), for $p = 1$, with a multiplicative triweight kernel. In order to analyze the effect of the bandwidth matrix in the test statistics, $T^{c,1}_{n,1}$ and $T^{c,2}_{n,1}$ are computed in a grid of (diagonal) matrices with both equal elements, $\mathbf{H} = \text{diag}(h, h)$. In this case, values of $h = 0.25, 0.40, 0.55, 0.70, 0.85, 1.00, 1.15, 1.30$ are considered. Moreover, the same weight function w as in Section 5.2.4 is used here.

In practice, in order to implement the bootstrap Algorithm 3, a wrapped Gaussian spatial process model is employed in Step 2. Following the proposal by Jona-Lasinio et al. (2012), the model fitting within a Bayesian framework is performed using a Markov Chain Monte Carlo method. Assuming a linear Gaussian spatial process of the form (4.1), to perform a Bayesian fit of the model, priors are needed for the model parameters. The authors suggest a normal prior for μ, a truncated inverse gamma prior for σ^2 and a uniform prior (with support allowing small ranges up to ranges

a bit larger than the maximum distance over the region) for the decay parameter $3/a_e$. More specifically, the prior of μ is a Gaussian distribution with zero mean and variance one. For σ^2, we choose an inverse Gamma $\text{IG}(a_\sigma, b_\sigma)$ with $a_\sigma = 2$ and $b_\sigma = 1$, then the mean is $b_\sigma/(a_\sigma - 1) = 1$. For the decay parameter, the uniform distribution on the interval $[3/a_e - 1, 3/a_e + 1)$ is considered when $a_e = 0, 1$, and on the interval $[3/a_e - 0.5, 3/a_e + 0.5)$ when $a_e = 0.3, 0.6$. The parameters are updated using a Metropolis–Hastings algorithm (Hastings, 1970). For further details on the wrapped Gaussian spatial model fitting we refer to Jona-Lasinio et al. (2012). The means of the posterior parameter estimates are considered in Step 3 of Algorithm 3. Notice that, in this case, the circular spatially correlated errors are generated from a wrapped Gaussian spatial process and, in Step 2 of Algorithm 3, a wrapped Gaussian spatial process model is employed for model fitting. This model only allows symmetric marginal distributions, therefore, if the errors were drawn by using other procedure, such as a projected Gaussian spatial processes (with asymmetric marginals), it would be more convenient to use an alternative approach. The impact of using a model that differs from the one used to generate the errors is briefly discussed in Section 6.4.

Effect of sample size. Table 5.13 shows the proportions of rejections of the null hypothesis for different sample sizes and $\alpha = 0.05$, when using $T_{n,1}^{c,1}$. Under the null hypothesis ($c = 0$), it can be observed that the test has an acceptable performance using both bootstrap approaches PSCB and NPSCB. Proportions of rejections are similar to the theoretical level considered, namely $\alpha = 0.05$. However, these proportions clearly depend on the value of the bandwidth h. For alternative assumptions ($c = 1$ and $c = 2$), the performance of the test is satisfactory. As expected, the power of the test is larger when the value of c is also larger. A slightly better performance of the test is obtained when considering the test statistic $T_{n,1}^{c,2}$. In this case, the proportions of rejections of the null hypothesis are presented in Table 5.14.

Effect of a_e. Results for $n = 400$ and different spatial dependence degrees ($a_e = 0.1, 0.3, 0.6$) are shown in Table 5.15, when using $T_{n,1}^{c,1}$. PSCB and NPSCB approaches provide good results for both the null and the alternative hypotheses. As expected, the power of the test is larger when the dependence structure is weaker. In these scenarios, results considering the test statistic $T_{n,1}^{c,2}$ are summarized in Table 5.16.

c	n	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	100	PSCB	0.030	0.032	0.028	0.032	0.028	0.024	0.026	0.030
		NPSCB	0.058	0.044	0.054	0.050	0.052	0.046	0.044	0.044
	225	PSCB	0.032	0.028	0.034	0.028	0.030	0.028	0.024	0.020
		NPSCB	0.044	0.040	0.050	0.038	0.040	0.040	0.040	0.038
	400	PSCB	0.030	0.028	0.036	0.036	0.036	0.030	0.034	0.034
		NPSCB	0.042	0.042	0.042	0.044	0.046	0.040	0.036	0.034
1	100	PSCB	0.088	0.192	0.238	0.248	0.246	0.236	0.216	0.208
		NPSCB	0.172	0.246	0.278	0.292	0.286	0.266	0.252	0.238
	225	PSCB	0.146	0.216	0.228	0.222	0.214	0.206	0.198	0.192
		NPSCB	0.188	0.244	0.252	0.244	0.242	0.228	0.220	0.216
	400	PSCB	0.166	0.210	0.204	0.212	0.212	0.198	0.190	0.180
		NPSCB	0.222	0.254	0.264	0.262	0.242	0.232	0.216	0.210
2	100	PSCB	0.670	0.702	0.722	0.718	0.696	0.686	0.668	0.656
		NPSCB	0.732	0.744	0.748	0.746	0.724	0.716	0.700	0.690
	225	PSCB	0.602	0.628	0.620	0.612	0.598	0.582	0.552	0.550
		NPSCB	0.658	0.670	0.656	0.644	0.628	0.614	0.600	0.590
	400	PSCB	0.562	0.576	0.570	0.560	0.548	0.520	0.514	0.506
		NPSCB	0.610	0.604	0.596	0.586	0.572	0.558	0.538	0.528

Table 5.13: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different sample sizes. Model parameters: $\sigma^2 = 1$ and $a_e = 0.3$. The test statistic $T_{n,1}^{c,1}$ is employed, and the Algorithm 3 is used. Significance level: $\alpha = 0.05$.

c	n	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	100	PSCB	0.030	0.032	0.028	0.030	0.028	0.024	0.030	0.032
		NPSCB	0.058	0.044	0.056	0.050	0.046	0.046	0.044	0.044
	225	PSCB	0.032	0.028	0.028	0.028	0.026	0.030	0.024	0.022
		NPSCB	0.044	0.044	0.048	0.040	0.034	0.038	0.040	0.038
	400	PSCB	0.030	0.028	0.034	0.034	0.034	0.028	0.034	0.036
		NPSCB	0.040	0.040	0.042	0.042	0.044	0.036	0.040	0.040
1	100	PSCB	0.120	0.178	0.254	0.280	0.262	0.236	0.212	0.202
		NPSCB	0.180	0.272	0.290	0.308	0.300	0.288	0.282	0.278
	225	PSCB	0.146	0.220	0.238	0.230	0.224	0.218	0.212	0.206
		NPSCB	0.188	0.248	0.258	0.246	0.244	0.236	0.236	0.232
	400	PSCB	0.168	0.214	0.222	0.220	0.224	0.220	0.222	0.224
		NPSCB	0.226	0.262	0.274	0.264	0.250	0.248	0.258	0.260
2	100	PSCB	0.674	0.710	0.732	0.732	0.716	0.700	0.690	0.686
		NPSCB	0.738	0.758	0.762	0.762	0.754	0.740	0.730	0.722
	225	PSCB	0.608	0.644	0.642	0.626	0.618	0.614	0.610	0.604
		NPSCB	0.670	0.686	0.678	0.664	0.652	0.646	0.642	0.640
	400	PSCB	0.566	0.590	0.582	0.580	0.582	0.570	0.572	0.566
		NPSCB	0.622	0.624	0.618	0.614	0.610	0.608	0.602	0.594

Table 5.14: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}_{2,\beta}$ with different sample sizes. Model parameters: $\sigma^2 = 1$ and $a_e = 0.3$. The test statistic $T_{n,1}^{c,2}$ is employed, and the Algorithm 3 is used. Significance level: $\alpha = 0.05$.

c	a_e	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	0.1	PSCB	0.036	0.032	0.044	0.040	0.042	0.042	0.038	0.034
		NPSCB	0.036	0.038	0.048	0.040	0.048	0.044	0.038	0.034
	0.3	PSCB	0.030	0.028	0.036	0.036	0.036	0.030	0.034	0.034
		NPSCB	0.042	0.042	0.042	0.044	0.046	0.040	0.036	0.034
	0.6	PSCB	0.032	0.026	0.032	0.026	0.024	0.024	0.016	0.016
		NPSCB	0.048	0.042	0.046	0.034	0.032	0.030	0.022	0.018
1	0.1	PSCB	0.654	0.650	0.636	0.600	0.572	0.554	0.540	0.528
		NPSCB	0.706	0.692	0.670	0.644	0.614	0.590	0.564	0.550
	0.3	PSCB	0.166	0.210	0.204	0.212	0.212	0.198	0.190	0.180
		NPSCB	0.222	0.254	0.264	0.262	0.242	0.232	0.216	0.210
	0.6	PSCB	0.088	0.102	0.114	0.116	0.104	0.104	0.100	0.088
		NPSCB	0.150	0.156	0.164	0.160	0.144	0.142	0.132	0.124
2	0.1	PSCB	0.800	0.798	0.792	0.766	0.734	0.720	0.682	0.674
		NPSCB	0.838	0.834	0.822	0.808	0.786	0.748	0.726	0.702
	0.3	PSCB	0.562	0.576	0.570	0.560	0.548	0.520	0.514	0.506
		NPSCB	0.610	0.604	0.596	0.586	0.572	0.558	0.538	0.528
	0.6	PSCB	0.286	0.336	0.368	0.364	0.344	0.320	0.298	0.276
		NPSCB	0.526	0.542	0.534	0.524	0.508	0.492	0.466	0.450

Table 5.15: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different values of a_e. Model parameters: $\sigma^2 = 1$ and $n = 400$. The test statistic $T^{c,1}_{n,1}$ is employed, and the Algorithm 3 is used. Significance level: $\alpha = 0.05$.

c	a_e	Method	$h = 0.25$	$h = 0.40$	$h = 0.55$	$h = 0.70$	$h = 0.85$	$h = 1.00$	$h = 1.15$	$h = 1.30$
0	0.1	PSCB	0.036	0.030	0.042	0.040	0.044	0.042	0.038	0.038
		NPSCB	0.036	0.040	0.048	0.040	0.048	0.042	0.038	0.038
	0.3	PSCB	0.030	0.028	0.034	0.034	0.034	0.028	0.034	0.036
		NPSCB	0.040	0.040	0.042	0.042	0.044	0.036	0.040	0.040
	0.6	PSCB	0.034	0.026	0.028	0.026	0.026	0.026	0.018	0.018
		NPSCB	0.048	0.038	0.044	0.036	0.032	0.030	0.026	0.022
1	0.1	PSCB	0.658	0.668	0.658	0.626	0.602	0.610	0.616	0.612
		NPSCB	0.718	0.706	0.692	0.670	0.668	0.650	0.660	0.662
	0.3	PSCB	0.168	0.214	0.222	0.220	0.224	0.220	0.222	0.224
		NPSCB	0.226	0.262	0.274	0.264	0.250	0.248	0.258	0.260
	0.6	PSCB	0.092	0.108	0.114	0.120	0.114	0.114	0.108	0.104
		NPSCB	0.152	0.158	0.166	0.166	0.160	0.158	0.154	0.152
2	0.1	PSCB	0.816	0.824	0.820	0.810	0.808	0.806	0.804	0.806
		NPSCB	0.856	0.860	0.860	0.850	0.846	0.846	0.844	0.840
	0.3	PSCB	0.566	0.590	0.582	0.580	0.582	0.570	0.572	0.566
		NPSCB	0.622	0.624	0.618	0.614	0.610	0.608	0.602	0.594
	0.6	PSCB	0.294	0.352	0.374	0.384	0.378	0.364	0.344	0.334
		NPSCB	0.532	0.548	0.546	0.538	0.534	0.532	0.522	0.510

Table 5.16: Proportions of rejections of the null hypothesis for the parametric family $\mathcal{M}^c_{2,\beta}$ with different values of a_e. Model parameters: $\sigma^2 = 1$ and $n = 400$. The test statistic $T^{c,2}_{n,1}$ is employed, and the Algorithm 3 is used. Significance level: $\alpha = 0.05$.

Chapter 6

Conclusions and discussion

This dissertation has focused on studying some inference problems for regression models with Euclidean and circular data, considering scenarios where the data are independent and also spatially correlated.

The first aim was to propose and analyze goodness-of-fit tests for multiple regression models with spatially correlated errors. These contents are included in Chapter 2. The second goal focused on proposing nonparametric approaches to estimate the regression function in a linear-circular regression model (circular response and Euclidean covariates). The results on this issue are collected in Chapter 3. The third objective is centered on extending the previous proposals, but assuming that the circular errors of the regression model are spatially correlated. The developments of this point are provided in Chapter 4. The fourth aim was to introduce goodness-of-fit tests for multiple linear-circular regression models with independent and spatially correlated data. The empirical analysis of the proposed procedures is presented in Chapter 5. In what follows, some final comments and discussion on each chapter are presented.

6.1 Goodness-of-fit tests for parametric regression models with spatially correlated errors

A goodness-of-fit test for a parametric regression model with spatially correlated errors was presented in Chapter 2. This test is based on a L_2-distance between a

smoothed version of a parametric fit and a nonparametric estimator of the regression function. A least squares procedure has been considered as a parametric approach, given its efficiency, but other procedures such as maximum likelihood methods, could be also used, as long as a $\sqrt{n}$-consistency property is satisfied. In this case, it should be noted that both the regression function and the dependence structure of the errors are jointly estimated, but usually restricted to a (multivariate) Gaussian distribution of the process realization. In both cases (least squares and maximum likelihood), a parametric form for the correlation is considered. Without being the target function and viewing spatial correlation as a nuisance (that should certainly be accounted for, but it is not of primary interest), it is expected that the proposed goodness-of-fit test has a good performance even when the correlation is misspecified as long as it can be reasonably well approximated. Testing approaches as those introduced by Maglione and Diblasi (2004) can be useful for this task. For the nonparametric fit of the proposed test statistic, the Nadaraya–Watson and local linear estimators were considered. However, the procedure could be generalized and studied for higher polynomial p degrees. Other kernel estimators such as Priestley–Chao estimator could be also used. This estimator has a very simple expression and can be applied to multidimensional problems. This is not the case of the Gasser-Muller estimator, which is difficult to extend to the multidimensional case because its calculation involves sorting and taking middle points in the design space.

The asymptotic distribution of the test, under the null and under local alternatives, was derived considering the assumption of increasing-domain spatial asymptotics. For practical implementation, due to the slow convergence to the limit distribution, resampling methods were used to calibrate the test. Specifically, three bootstrap procedures were designed and applied in practice: PB, NPB and CNPB. The NPB and CNPB resampling methods avoid model selection and, therefore, prevents against misspecification problems in the estimation of the regression function and dependence structure, unlike the PB approach. The CNPB also corrects the bias induced by the use of the residuals in the approximation of the dependence structure, using an iterative method, providing good results of the test under the null and alternative hypotheses. On the one hand, as it was pointed out by Fernández-Casal and Francisco-Fernández (2014), a similar tool for bias adjustment could be included

in the parametric semivariogram estimation in the PB approach (see Davison and Hinkley, 1997). However, this way of proceeding would not avoid the misspecification problem in the parametric estimation of both the semivariogram and the regression function. On the other hand, given that a composite hypothesis is tested in (1.34), a double bootstrap (Beran, 1987, 1988; Hall, 1986) could be included in order to improve the effectiveness of bias correction in the PB approach.

As usual in this type of problems, the performance of the goodness-of-fit test has been explored in a grid of different bandwidths to check how it is affected by the bandwidth choice. Simulations were carried out when considering regression model (1.1) with different regression functions and random errors drawn from Gaussian or non-Gaussian processes. No major differences (in terms of performance of the test) have been found if Gaussian or non-Gaussian processes are used to generate the errors. In the vast majority of scenarios considered in the simulation study, results obtained by the CNPB improve those achieved by PB and NPB. The use of non-scalar bandwidths has not provided better results for CNPB. The PB proposal works properly for calibration, but it shows a limited capacity to detect alternatives. On the other hand, although similar resampling methods to NPB have given good results when are employed in goodness-of-fit tests considering regression models with independent and univariate data, this is not the case in the spatial framework. In this setting, the proportions of rejections under the null hypothesis are very large compared with the significance level considered, due to the underestimation of the variability of the process.

The three resampling approaches compared in this research are based on computing the residuals from a pilot fit, estimating the corresponding covariance matrix of the errors and, finally, using a Cholesky decomposition to approximate a vector of *independent* errors to generate bootstrap resamples. Other resampling procedures, such as the block bootstrap (see Lahiri, 2003), could be used to calibrate the test. This method requires an appropriate partition of the observation region, unlike parametric and nonparametric bootstrap-based methods. In addition, block bootstrap-based approaches present difficulties when the interest is focused on estimating the second-order structure (dependence) of the process, which is often necessary to estimate properly the large-scale variability. These procedures fail to reproduce the

variability of the process, thus leading to an underestimation of the semivariogram, possibly caused by the selection of the blocks. In Castillo-Páez et al. (2019), parametric, corrected nonparametric and block bootstrap mechanisms were compared by checking their performance in the approximation of the bias and the variance of two variogram estimators. For inference on spatial processes and, particularly, on dependence structure estimation, the authors recommend the use of corrected nonparametric bootstrap methods. For these reasons, block bootstrap approaches were not employed in the present research.

The consistency of the proposed bootstrap algorithms could be analyzed, for instance, by an *imitation* procedure (Shao and Tu, 2012). It would be also interesting to consider a wild bootstrap approach to generate the independent errors to be multiplied by the matrix derived from the Cholesky decomposition in Algorithm 1.

As pointed out before, the type of tests considered in Chapter 2 are smooth-based tests taking an L_2-distance for constructing the test statistics. Given hat these procedures rely on the use of smooth regression curves, obtained by local-constant and local-linear fits, the tests somehow inherit two of the shortcomings of kernel methods. On the one hand, a bandwidth (a matrix, in this case) must be selected and, on the other hand, the curse-of-dimensionality hampers the practical use of this approach for higher dimensions. An alternative route that may be explored for constructing goodness-of-fit tests in this context, avoiding the selection of the bandwidth, may be found using the ideas employed in correlation-distance based tests (see Székely et al., 2007). Actually, this type of approaches are not apparently affected by the increase of dimensionality, so it will facilitate the application of the methods in other settings with a higher number of covariates.

Although a homoscedastic regression model has been considered in Chapter 2, under suitable assumptions, the asymptotic results of the proposed test statistic could be also derived for certain heteroscedastic regression models. In such a context, the bootstrap methods to calibrate the test, described in Section 2.3, could be also modified, using an appropriate route to estimate the dependence of the model. To do this, the nonparametric approach described by Fernández-Casal et al. (2017) could be used. Note that in that case, due to heteroscedasticity, the use of a wild bootstrap procedure in the different resampling processes could be more convenient. The design

of this type of resampling approach in this context is, indeed, an interesting issue for a future research.

In practice, the numerical studies performed in Chapter 2 (and in the rest of chapters of this dissertation) were run in an Intel Core i7-9700K at 3.60Ghz. The procedures used in the simulation study and in the illustration with real data were implemented in the statistical environment `R` (R Development Core Team, 2020), using functions included in the `npsp` and `geoR` packages (Fernández-Casal, 2019; Ribeiro and Diggle, 2020) to estimate the variogram and the spatial regression functions. In particular, the bias correction in CNPB bootstrap algorithm is implemented in the function `np.svariso.corr` of the `npsp` `R` package. The computing time for running the whole testing procedure (simulate a sample, compute the test statistics in a range of bandwidths and apply the bootstrap methods considering $B = 500$ replications) for a sample of size $n = 100, 225$ and 400 is around 2, 8 and 27 seconds, respectively, no matter the nonparametric estimator (Nadaraya–Watson or local linear) used. However, PB seems to be slightly more computationally expensive than NPB and CNPB.

6.2 Nonparametric regression estimation for a circular response and an $\mathbb{R}^d$-valued covariate

Nonparametric regression estimation for circular response and $\mathbb{R}^d$-valued covariate was studied in Chapter 3. The proposal considers kernel-based approaches, with special attention on Nadaraya–Watson- and local linear-type estimators in general dimension, and for higher order polynomials in the one-dimensional case. Asymptotic conditional bias and variance were derived and the performance of the estimators was assessed in a simulation study.

One of the advantages of the proposed procedure (if the bandwidth matrix is appropriately chosen, and a suitable bandwidth selector is also provided in that chapter) is that it relaxes parametric assumptions, and consequently enables one to explore and model the data more flexibly, avoiding misspecification problems. Moreover, this estimator can be employed as a first attempt to explore if a certain parametric family is appropriate or not to model the data.

For practical implementation, the selection of a d-dimensional bandwidth matrix is required. In the regression Euclidean context, the bandwidth selection problem has been widely addressed in the last decades (see, for example, Köhler et al., 2014, where a review on bandwidth selection methods for kernel regression was provided). More related to the topic of Chapter 3, a rule-of-thumb and a bandwidth rule for selecting scalar or diagonal bandwidth matrices for multivariate local linear regression with real-valued response and $\mathbb{R}^d$-valued covariate was derived in Yang and Tschernig (1999). Also in that context, in González-Manteiga et al. (2004), a bootstrap method to estimate the mean squared error and the smoothing parameter for the multidimensional regression local linear estimator was proposed. However, in the framework of nonparametric regression methods for circular variables, the research on bandwidth selection is very scarce or non-existent.

The practical results obtained in Chapter 3 were derived with a cross-validation bandwidth given that, up to the knowledge of the author, there are no other bandwidth selectors available in this context. Cross-validation techniques have the drawback of being unable to provide satisfactory results in a reasonable time for very large sample sizes due to its computational complexity. To overcome this problem, bagging cross-validation bandwidths studied for density and regression estimation with Euclidean data (Hall and Robinson, 2009) could be adapted to this context. Additionally, note that even though cross-validation bandwidths present appealing theoretical properties, in practice, their computation could present certain difficulties in a multidimensional framework.

The design of alternative procedures to select the bandwidth matrix for the estimators studied in this framework based, for example, on bootstrap methods are indeed of great interest. This problem is out of the scope of this thesis, but it is an interesting topic of research for a future study. On the other hand, it should be noted that just global bandwidths are considered in the numerical studies of this dissertation. A limitation derived from the use of global bandwidths is that the corresponding nonparametric estimators may provide spurious estimates in areas with sparse observations. In such regions, the number of observations within the neighborhood determined by the bandwidth may be too small, producing unstable estimates. A way to overcome this problem is employing local bandwidths that automatically

adapt to the number of observations near the grid points where calculating the estimates. The problem of using local bandwidths is that a high computing time would be required, specially if the sample size is very large.

Once the problem of including a $\mathbb{R}^d$-valued covariate for explaining the behavior of a circular response is solved, it seems natural to think about the consideration of covariates of different nature. Since the proposed estimator was constructed by considering the atan2 of the smooth estimators of the regression functions for the sine and cosine components of the response, an adaptation of our proposal for different types of covariates implies the use of suitable weights. For instance, if a spherical (circular, as a particular case) or a mixture of spherical and real-valued covariates are considered to influence a circular response, weights for estimating the sine and cosine components could be constructed following the ideas in García-Portugués et al. (2013) for cylindrical density estimation. If a categorical covariate is included in the model, a similar approach to the one in Racine and Li (2004) or in Li and Racine (2004) could be also followed. In all these cases, bandwidth matrices should be selected, and cross-validation techniques could be applied.

The results obtained in Theorems 3.3 and 3.4 can be extended to an arbitrary dimension d of the space of the covariates by using the asymptotic properties for $\hat{m}_{j,\mathbf{H}}(\mathbf{x};p)$, for $p=2,3$, provided in Gu et al. (2015). The authors considered the leading term of biases and variances of multivariate local polynomial estimators of general order p. Results on the asymptotic distribution of multivariate local polynomial estimators are also provided in Gu et al. (2015). The joint asymptotic normality of $\hat{m}_{1,\mathbf{H}}(\mathbf{x};p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};p)$ can be used to derive, via the delta-method, the asymptotic distribution of statistics which can be expressed in terms of $\hat{m}_{1,\mathbf{H}}(\mathbf{x};p)$ and $\hat{m}_{2,\mathbf{H}}(\mathbf{x};p)$. For example, a suitable adaptation of Proposition 3.1 of Jammalamadaka and SenGupta (2001) can be used to derive the limiting distribution of the tangent of $\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)$.

The asymptotic normality of the proposed estimators would allow the construction of pointwise confidence intervals for the circular regression function m^c at $\mathbf{x}\in\mathcal{D}$. Notice that both bias and variance of the nonparametric circular regression estimators $m^c_{\mathbf{H}}(\mathbf{x};p)$ depend on unknown functions, therefore, in order to make use of the asymptotic result, they should be estimated. The coverage rates will be affected

by the bias of the circular regression estimators $m^c_{\mathbf{H}}(\mathbf{x}; p)$. Alternatively, confidence intervals for the circular regression function m^c could be also obtained by using bootstrap procedures (Hall, 1992).

In practice, the simulations were implemented in the statistical environment `R` (R Development Core Team, 2020), using functions included in the `npsp` and `circular` packages (Fernández-Casal, 2019; Lund et al., 2020). The real data application was performed in MATLAB software (www.mathworks.com). The computing time for running the procedure (simulate a sample, select the bandwidth matrix, compute the circular nonparametric estimator and evaluate the CASE) for just one sample of size of 225 is less than 2 seconds.

6.3 Nonparametric regression estimation for a circular response and an $\mathbb{R}^d$-valued covariate with spatially correlated errors

In the scenario of Chapter 3, data generated from the regression model are assumed to be independent, and nonparametric circular regression estimators were proposed. However, in many practical situations, the independence assumption does not seem reasonable (e.g. data area collected over time or space). The simple construction scheme behind the proposed class of estimators makes possible to easily obtain asymptotic properties in more general frameworks. Nonparametric regression estimators in a model with circular response and real-valued covariates in the presence of spatial correlation were studied in Chapter 4. The estimators proposed in Chapter 3 were also considered in this framework. The asymptotic conditional bias of the kernel-type proposed estimators is the same as that obtained for independent data. However, the asymptotic conditional variance depends on the spatial correlation.

Although there is a substantial literature on modeling circular data with spatial dependence by introducing and formulating spatial processes for circular data, such as wrapped and projected Gaussian spatial processes (Jona-Lasinio et al., 2012; Wang and Gelfand, 2014), our proposal follows a different perspective. We consider an appropriate linear-circular regression model for spatially correlated data and estimate

nonparametrically the corresponding circular spatial trend. This is an alternative to model circular data at different spatial locations. It should be noted that no other direct competitor, up to the knowledge of the author, has yet been proposed following these ideas, neither from a parametric nor from a nonparametric approach.

Regarding the bandwidth matrix needed to compute the circular regression estimators given in (3.14), it can be selected by leave-one-out cross-validation, but this matrix is not necessarily a good one for spatially correlated data, given that $\mathbb{E}[\mathrm{CV}(\mathbf{H})]$ (and consequently $\mathbb{E}[\mathrm{CV}^c(\mathbf{H})]$) is severely affected by the correlation (Liu, 2001; Opsomer et al., 2001). In this context, it is advisable to consider other bandwidth selection criteria that take the spatial dependence structure into account. In the practical results of this chapter, we also considered a modified cross-validation method suitably adjusted for the presence of spatial correlation, which considers to "leave N_i observations out". The idea of modifying the selection criterion in this manner is not new. An example of such adjustment is the "leave $2l+1$ out" cross-validation approach (Hart and Vieu, 1990). It should be noted that in the context of that paper, ordinary cross-validation criterion worked reasonably well for moderate correlation structures. Further, for mid to low values of the parameter which controls the dependence structure, it was obtained a very small improvement when using the modified cross-validation method. This was not the case for the spatial framework considered in this thesis, where results were considerably better when the modified cross-validation criterion (MCV^c) was used, compared to those obtained when simple cross-validation (CV^c) was employed. The importance of the dependence structure becomes more relevant as the dimension increases and, probably, for that reason, in the context of this thesis a clearer improvement is observed when the modified cross-validation criterion is used, no matter the dependence degree considered. Note that in the case of the marine currents in the Adriatic Sea, a suitable parameter l in MCV_l^c, controlling the number of observations left out, has been selected by minimizing the prediction error given in (4.17). An interesting point would be to design a fully automatic procedure to compute the optimal radius in the MCV^c method. This approach should account for the spatial correlation of the covariates and, although it is out of the scope of this dissertation, it would be an interesting topic of further research.

Alternatively to the cross-validation methods previously described, the bandwidth could be selected as follows. First, as described in Chapter 4, the whole sample is split in two parts, a training and a testing samples. First, as described in Section 4.6, the whole sample is split in two parts, a training and a testing samples. Then, using the overall dataset and a pilot bandwidth matrix, the nonparametric estimator is computed at each training data point. From this estimation, and using a wrapped (Jona-Lasinio et al., 2012) or a projected Gaussian spatial process (Wang and Gelfand, 2014), the residuals can be modeled, obtaining predictions at the testing locations. Finally, the bandwidth matrix can be selected by minimizing the corresponding prediction error computed with the testing sample. Notice that a drawback of this approach is a possible misspecification of the fit model. The modeling based on wrapped Gaussian spatial processes allows only symmetric marginal distributions, while projected Gaussian processes have marginal distributions that can be asymmetric, possibly bimodal.

In the current setting, real-valued covariates for explaining the behavior of a circular response in presence of spatial correlation are considered. However, it might be the case that other types of covariates, such as other circular, or more generally, spherical covariates, may influence the circular response. For these more complex scenarios, there is a substantial research on modeling and on analyzing inference approaches for random fields on spheres as well as on spheres across time. For instance, Porcu et al. (2016) developed cross-covariance functions of the great circle distances on the sphere. Alegría et al. (2019) proposed a flexible parametric family of matrix-valued covariance functions. To overcome the problem of generating samples from random fields, Emery and Porcu (2019) introduced an algorithm to generate isotropic vector-valued Gaussian random fields defined over the unit two-dimensional sphere embedded in the three-dimensional Euclidean space. Some of these approaches could be incorporated in model (4.5), and although these extensions are also out of the scope of this thesis, they can be the focus of future works.

In practice, the simulations were implemented in the statistical environment `R` (R Development Core Team, 2020), using functions included in the `npsp` and `CircSpaceTime` packages (Fernández-Casal, 2019; Jona-Lasinio et al., 2019). The real data application was performed in MATLAB software (www.mathworks.com).

The computing time for running the whole procedure for one sample of size of 225 is around 2 seconds, no matter the bandwidth matrix selection method employed and regardless of the estimator (Nadaraya–Watson or local linear) used. However, it should be noted that the computing time for obtaining a bandwidth matrix with the MCV criterion increases with l. In addition, the projection approach for the circular errors generation seems to be slightly more computationally expensive than the wrapping method.

6.4 Goodness-of-fit tests for parametric regression models with a circular response and an $\mathbb{R}^d$-valued covariate

Testing procedures for assessing a parametric circular regression model (with a circular response and an $\mathbb{R}^d$-valued covariate) were proposed and empirically analyzed in Chapter 5 for independent and for spatially correlated data. The proposed test statistics were constructed by measuring a circular distance between a (non-smoothed or smoothed) parametric fit and a nonparametric estimator of the circular regression function. For the parametric approach, taking into account that the classical least squares regression method is not appropriate when the response variable is of circular nature, a circular analog can be used (Fisher and Lee, 1992; Lund, 1999). Alternative parametric fitting approaches, such as maximum likelihood methods, could be also used (Presnell et al., 1998). Regarding the nonparametric fit, local polynomial type estimators, given in (3.14), were considered in the test statistics.

Although the calculation of the asymptotic distribution of the tests, under the null and under local alternatives, is out of the scope of this thesis, its derivation can follow from using a Taylor approximation of the function $1-\cos(\Theta)$ by $\Theta^2/2$, for $\Theta \in [0, 2\pi)$ (Kim and SenGupta, 2017). Using this approach, the expressions $1-\cos[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)-m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{x})]$ and $1-\cos[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)-\hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};p)]$ in the test statistics $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$, given in (5.4) and in (5.5), respectively, can be approximated by $1/2[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)-m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{x})]^2$ and $1/2[\hat{m}^c_{\mathbf{H}}(\mathbf{x};p)-\hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x};p)]^2$, respectively. Consequently, $T^{c,1}_{n,p}$ and $T^{c,2}_{n,p}$ can be approximated by test statistics similar to $T_{n,p}$, given in (2.1), for regression

models with Euclidean response and covariates. Notice that the regression estimators involved in the test statistics $T_{n,p}^{c,1}$ and $T_{n,p}^{c,2}$ have more complicated expressions than those in $T_{n,p}$. Therefore, as intuition suggests, it will be more difficult to calculate the close expressions of their asymptotic distributions.

For practical implementation, bootstrap resampling methods were used to calibrate the tests. For independent data, two procedures have been designed and compared: PCB and NPCB. Both methods are based on computing the residuals and generating independent bootstrap resamples. The main difference between them is the mechanism employed to obtain the residuals. In PCB, the residuals come from the parametric regression estimator. Alternatively, in NPCB, the residuals are obtained from the nonparametric regression estimator. For dependent data, in order to imitate the distribution of the (spatially correlated) errors, new bootstrap procedures were proposed: PSCB and NPSCB. Again, the main difference between both approaches is how the residuals are obtained. In the case of the PSCB, the residuals come from the parametric fit, whereas in NPSCB, the residuals are obtained from the nonparametric estimator. In practice, in order to implement the procedures, a wrapped Gaussian spatial process model (Jona-Lasinio et al., 2012) was fit to the residuals to mimic the dependence structure. This wrapped Gaussian spatial process model was fit within a Bayesian framework, therefore, some prior parameter values must be provided to use the Markov Chain Monte Carlo model fitting. For further details on wrapped Gaussian model fitting, we refer to Jona-Lasinio et al. (2012). Alternatively, other spatial-circular process models, such as asymmetric wrapped Gaussian spatial processes (Mastrantonio et al., 2016) or projected Gaussian spatial processes (Wang and Gelfand, 2014) could be employed to model the residuals, and thus try to imitate the dependence structure of the errors. The impact of using a model that differs from the one used to generate the errors was explored. For instance, if errors are drawn from a wrapped Gaussian spatial process, and a projected Gaussian spatial process if fit, the procedure works fairly well, for appropriate prior parameter values, since this model is highly flexible. Conversely, although the wrapping approach gives results that are really easy to interpret, a limitation derived from the use of a wrapped Gaussian spatial process is that such modeling allows only symmetric marginal distributions. If the errors are generated from an (asymmetric)

projected Gaussian spatial process and a wrapped Gaussian spatial process is fit, the results obtained using the testing procedure are not entirely satisfactory.

Once the model is fit, error bootstrap samples are generated from it. These errors bootstrap samples could be also employed to design a parametric iterative least squares estimator, accounting for the possible spatial dependence structure, that could be used in the tests for spatially correlated data (instead of the parametric estimator given in Section 5.2.1). Specifically, using the error bootstrap samples, the variance-covariance matrix of the circular errors can be approximated. Then, applying a Cholesky decomposition of this matrix, the original circular responses and the $\mathbb{R}^d$-valued covariate are transformed, as it is done in the generalized least squares method. Finally, the parameter estimate is obtained applying (5.2) to the transformed observations. Obviously, this algorithm could be applied iteratively. Although, we have not applied this method in practice, we do not believe that it provides great improvements over using the circular least squares method described in Section 5.2.1, even though the data are indeed dependent. The possible benefits of taking the correlation of the data into account could be offset by the difficulty of adequately estimating the variance-covariance matrix of the circular errors.

For independent data, in the majority of scenarios considered in the simulation study, results obtained with NPCB improved those achieved by PCB, especially, for alternative assumptions. Moreover, a better behavior is observed when $T_{n,1}^{c,2}$, given in (5.5), is employed. For spatially correlated data, it is obtained that both tests do not work properly under the null hypothesis, when using PCB and NPCB, designed for independence. Regarding PSCB and NPSCB, the use of the nonparametric residuals in the bootstrap procedure provides the best results. As expected, the power of the tests is larger when the spatial dependence structure is weaker. More satisfactory results are achieved when $T_{n,1}^{c,2}$ is used. In both frameworks (independent and spatially correlated data), the proportions of rejections of the null hypothesis clearly depends on the bandwidth matrix considered. The whole simulation study (for independent and for spatially correlated data) was repeated when using the Nadaraya–Watson-type estimator for the nonparametric fits, given in (5.4) and in (5.5), for $p = 0$. For independent data, the procedures work fairly well when PCB is employed, while NPCB provides quite poor results. Similar conclusions can be derived for spatially

correlated data, where the tests have an acceptable performance using PSCB, while results obtained when NPSCB approach is employed are generally not good. It seems that the tests statistics suffer from boundary problems induced by the use of the Nadaraya–Wtason-type estimator, while this issue is overcome employing the local linear-type estimator. When using the Nadaraya–Watson-type estimator, probably a modification of the weight functions w used in the simulation study is required to obtain better results. Although the test statistics $T_{n,p}^{c,1}$ and $T_{n,p}^{c,2}$, given in (5.4) and in (5.5), were presented and numerically analyzed only for $p = 0, 1$, they may be also defined considering a local polynomial-type estimator of a general order p. Nevertheless, significantly better results than the ones obtained for $p = 1$ (local linear case) are not expected.

The procedures used in the simulation study were implemented in the statistical environment `R` (R Development Core Team, 2020), using functions included in the `npsp`, `CircSpaceTime` and `Rfast` packages (Fernández-Casal, 2019; Jona-Lasinio et al., 2019; Papadakis et al., 2020). For regression models with a single real-valued covariate, the computing time for running the whole testing procedure (simulate a sample, compute the test statistics in a range of bandwidths and apply the bootstrap methods considering $B = 500$ replications) for a sample of size $n = 50, 100$ and 200 is around 2, 3 and 5 seconds, respectively, no matter the bootstrap method (PCB or NPCB) used to calibrate the test. For a bidimensional one, the computing times are around 4, 6 and 14 seconds, when $n = 100, 225$ and 400, respectively. As expected, considering a bidimensional covariate is more computationally expensive than using a single covariate. In the spatial framework, the computing time for running the testing procedure for a sample of size $n = 100, 225$ and 400 is around 3, 7 and 15, respectively, when using Algorithm 2. This time is slightly increased when Algorithm 3 is employed (4, 13 and 49 seconds, respectively).

Appendix A

Auxiliary results

This appendix contains some auxiliary results employed in the proofs of Chapters 2, 3 and 4 of this dissertation. Some helpful results involving kernel and correlation functions are provided in Section A.1. For a fixed design scheme, Riemann approximations of sums by integrals are provided in Section A.2. Finally, Section A.3 contains other helpful results.

A.1 Some results involving kernel and correlation functions

In this section, some results including kernel and correlation functions are given. In order to prove some of these results, notice that, using Chebychev's inequality,

$$\mathbf{X}_n = \mathbb{E}(\mathbf{X}_n) + \mathcal{O}_{\mathbb{P}}[\sqrt{\mathrm{Var}(\mathbf{X}_n)}],$$

being $\mathbf{X}_n$ a stochastic sequence such that each element has finite variance.

Lemma A.1 *Let*

$$\begin{aligned}
\tilde{k}_{1,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),\\
\tilde{k}_{2,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),\\
\tilde{k}_{3,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^{\mathrm{T}}g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),\\
\tilde{k}_{4,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K^2_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),\\
\tilde{k}_{5,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K^2_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),\\
\tilde{k}_{6,n}(\mathbf{x}) &= n^{-1}\textstyle\sum_{i=1}^{n} K^2_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})(\mathbf{X}_i-\mathbf{x})^{\mathrm{T}}g_1(\mathbf{X}_i)g_2(\mathbf{X}_i),
\end{aligned}$$

where g_1 and g_2 are bounded functions at $\mathbf{x}$, under assumptions(A1), (H2) *and* (K1), *if $\mathbf{x}$ is an interior point of the support of f, then,*

$$\tilde{k}_{1,n}(\mathbf{x}) = f(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x}) + o_{\mathbb{P}}(1), \tag{A.1}$$

$$\tilde{k}_{2,n}(\mathbf{x}) = \mu_2(K)\boldsymbol{\nabla} f(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x})\mathbf{H}^2 + o_{\mathbb{P}}(\mathbf{H}^2\mathbf{1}_d), \tag{A.2}$$

$$\tilde{k}_{3,n}(\mathbf{x}) = \mu_2(K)f(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x})\mathbf{H}^2 + o_{\mathbb{P}}(\mathbf{H}\mathbf{1}_{d\times d}\mathbf{H}), \tag{A.3}$$

$$\tilde{k}_{4,n}(\mathbf{x}) = |\mathbf{H}|^{-1}R(K)f(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x}) + o_{\mathbb{P}}\left(|\mathbf{H}|^{-1}\right), \tag{A.4}$$

$$\tilde{k}_{5,n}(\mathbf{x}) = |\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_d), \tag{A.5}$$

$$\tilde{k}_{6,n}(\mathbf{x}) = |\mathbf{H}|^{-1}o_{\mathbb{P}}(\mathbf{1}_{d\times d}). \tag{A.6}$$

Proof of Lemma A.1 Firstly, expression (A.1) is proved. Notice that

$$\begin{aligned}
\mathbb{E}[\tilde{k}_{1,n}(\mathbf{x})] &= \int K_{\mathbf{H}}(\mathbf{u}-\mathbf{x})\, g_1(\mathbf{u})g_2(\mathbf{u})f(\mathbf{u})d\mathbf{u}\\
&= \int K(\mathbf{p})\, g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}\\
&= [g_1(\mathbf{x})+o(1)][g_2(\mathbf{x})+o(1)][f(\mathbf{x})+o(1)]\\
&= g_1(\mathbf{x})g_2(\mathbf{x})f(\mathbf{x})+o(1),\\
\mathbb{V}\mathrm{ar}[\tilde{k}_{1,n}(\mathbf{x})] &\leq n^{-1}\int K^2_{\mathbf{H}}(\mathbf{u}-\mathbf{x})\, g_1^2(\mathbf{u})g_2^2(\mathbf{u})f(\mathbf{u})d\mathbf{u}\\
&= \frac{1}{n|\mathbf{H}|}\int K^2(\mathbf{p})\, g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p})g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}\\
&= \frac{1}{n|\mathbf{H}|}R(K)[g_1(\mathbf{x})+o(1)]^2[g_2(\mathbf{x})+o(1)]^2[f(\mathbf{x})+o(1)].
\end{aligned}$$

Under (H2), and thus (H1), it follows that

$$\mathbb{V}\mathrm{ar}[\tilde{k}_{1,n}(\mathbf{x})] \;=\; o(1).$$

Now, expression (A.2) is proved. First,

$$\begin{aligned}
\mathbb{E}[\mathbf{H}^{-2}\tilde{k}_{2,n}(\mathbf{x})] &= \mathbf{H}^{-2}\int K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right) g_1(\mathbf{u})g_2(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \mathbf{H}^{-1}\int K\left(\mathbf{p}\right)\mathbf{p}g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= \mathbf{H}^{-1}\int K\left(\mathbf{p}\right)\mathbf{p}g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{p}) \\
&\quad \cdot[f(\mathbf{x})+\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\nabla}f(\mathbf{x})+\frac{1}{2}\mathbf{p}^{\mathrm{T}}\mathbf{H}\boldsymbol{\mathcal{H}}_f(\mathbf{x})\mathbf{H}\mathbf{p}]d\mathbf{p} \\
&= \mu_2(K)\boldsymbol{\nabla}f(\mathbf{x})[g_1(\mathbf{x})+o(1)][g_2(\mathbf{x})+o(1)] \\
&= \mu_2(K)\boldsymbol{\nabla}f(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x})+o(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\mathbf{H}^{-2}\tilde{k}_{2,n}(\mathbf{x})] &\leq n^{-1}\mathbf{H}^{-2}\int K_{\mathbf{H}}^2\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)^{\mathrm{T}} g_1^2(\mathbf{u})g_2^2(\mathbf{u}) \\
&\quad \cdot f(\mathbf{u})d\mathbf{u}\mathbf{H}^{-2} \\
&= \frac{1}{n|\mathbf{H}|}\mathbf{H}^{-2}\int K^2\left(\mathbf{p}\right)\mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H}g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p})g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p}) \\
&\quad \cdot f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}\mathbf{H}^{-2} \\
&= \frac{1}{n|\mathbf{H}|}\mathbf{H}^{-2}\mu_2(K)\boldsymbol{\nabla}f(\mathbf{x})[g_1(\mathbf{x})+o(1)]^2[g_2(\mathbf{x})+o(1)]^2 \\
&\quad \cdot[f(\mathbf{x})+o(1)].
\end{aligned}$$

Under (H2), it follows that $n|\mathbf{H}|\lambda_{\min}^2(\mathbf{H})\to\infty$, when $n\to\infty$, which is equivalent to $n^{-1}|\mathbf{H}|^{-1}\mathbf{H}^{-2}\to 0$, when $n\to\infty$, and therefore

$$\mathbb{V}\mathrm{ar}[\mathbf{H}^{-2}\tilde{k}_{2,n}(\mathbf{x})] \;=\; o(\mathbf{1}_d).$$

Regarding expression (A.3), it follows that

$$\begin{aligned}
\mathbb{E}[\mathbf{H}^{-2}\tilde{k}_{3,n}(\mathbf{x})] &= \mathbf{H}^{-2}\int K_{\mathbf{H}}\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)^{\mathrm{T}} g_1(\mathbf{u})g_2(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \mathbf{H}^{-2}\int K\left(\mathbf{p}\right)\mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H}g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= \mu_2(K)\mathbf{I}_d[g_1(\mathbf{x})+o(1)][g_2(\mathbf{x})+o(1)][f(\mathbf{x})+o(1)] \\
&= \mu_2(K)\mathbf{I}_d g_1(\mathbf{x})g_2(\mathbf{x})f(\mathbf{x})+o(1).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[\mathbf{H}^{-2}\tilde{k}_{3,n}(\mathbf{x})] &\leq n^{-1}\mathbf{H}^{-2}\int K_{\mathbf{H}}^2\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)^{\mathrm{T}}\left(\mathbf{u}-\mathbf{x}\right)\left(\mathbf{u}-\mathbf{x}\right)^{\mathrm{T}} \\
&\quad \cdot g_1^2(\mathbf{u})g_2^2(\mathbf{u})f(\mathbf{u})d\mathbf{u}\mathbf{H}^{-2} \\
&= \frac{1}{n|\mathbf{H}|}\mathbf{H}^{-2}\int K^2\left(\mathbf{p}\right)\mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H}\mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H}g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p})g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p}) \\
&\quad \cdot f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p}\mathbf{H}^{-2}.
\end{aligned}$$

Using (H2), and hence (H1), one gets that

$$\mathbb{V}\mathrm{ar}[\mathbf{H}^{-2}\tilde{k}_{3,n}(\mathbf{x})] = o(\mathbf{1}_{d\times d}).$$

Now, (A.4) is proved. Notice that

$$\begin{aligned}
\mathbb{E}[|\mathbf{H}|\tilde{k}_{4,n}(\mathbf{x})] &= |\mathbf{H}|\int K_{\mathbf{H}}^2\left(\mathbf{u}-\mathbf{x}\right)g_1(\mathbf{u})g_2(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \int K^2\left(\mathbf{p}\right)g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= R(K)[g_1(\mathbf{x})+o(1)][g_2(\mathbf{x})+o(1)][f(\mathbf{x})+o(1)] \\
&= R(K)g_1(\mathbf{x})g_2(\mathbf{x})f(\mathbf{x})+o(1), \\
\mathbb{V}\mathrm{ar}[|\mathbf{H}|\tilde{k}_{4,n}(\mathbf{x})] &\leq n^{-1}|\mathbf{H}|^2\int K_{\mathbf{H}}^4\left(\mathbf{u}-\mathbf{x}\right)g_1^2(\mathbf{u})g_2^2(\mathbf{u})f(\mathbf{u})d\mathbf{u} \\
&= \frac{1}{n|\mathbf{H}|}\int K^4\left(\mathbf{p}\right)g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p})g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p})f(\mathbf{x}+\mathbf{H}\mathbf{p})d\mathbf{p} \\
&= o\left(1\right).
\end{aligned}$$

As for expression (A.5), it follows that

$$\begin{aligned}\mathbb{E}[|\mathbf{H}|\tilde{k}_{5,n}(\mathbf{x})] &= |\mathbf{H}| \int K_{\mathbf{H}}^2 (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x}) g_1(\mathbf{u}) g_2(\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= \int K^2 (\mathbf{p}) \mathbf{H}\mathbf{p} g_1(\mathbf{x}+\mathbf{H}\mathbf{p}) g_2(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{p}) d\mathbf{p} \\ &= o(\mathbf{1}_d),\end{aligned}$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}[|\mathbf{H}|\tilde{k}_{5,n}(\mathbf{x})] &\leq n^{-1}|\mathbf{H}|^2 \int K_{\mathbf{H}}^4 (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x})^{\mathrm{T}} g_1^2(\mathbf{u}) g_2^2(\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{n|\mathbf{H}|} \int K^4 (\mathbf{p}) \mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H} g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p}) g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{p}) d\mathbf{p} \\ &= o(\mathbf{1}_d).\end{aligned}$$

Finally, in order to prove expression (A.6), note that

$$\begin{aligned}\mathbb{E}[|\mathbf{H}|\tilde{k}_{6,n}(\mathbf{x})] &= |\mathbf{H}| \int K_{\mathbf{H}}^2 (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x})^{\mathrm{T}} g_1(\mathbf{u}) g_2(\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= \int K^2 (\mathbf{p}) \mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H} g_1(\mathbf{x}+\mathbf{H}\mathbf{p}) g_2(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{p}) d\mathbf{p} \\ &= o(\mathbf{1}_{d\times d}),\end{aligned}$$

$$\begin{aligned}\mathbb{V}\mathrm{ar}[|\mathbf{H}|\tilde{k}_{6,n}(\mathbf{x})] &\leq n^{-1}|\mathbf{H}|^2 \int K_{\mathbf{H}}^4 (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x})^{\mathrm{T}} (\mathbf{u}-\mathbf{x}) (\mathbf{u}-\mathbf{x})^{\mathrm{T}} \\ &\quad \cdot g_1^2(\mathbf{u}) g_2^2(\mathbf{u}) f(\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{n|\mathbf{H}|} \int K^4 (\mathbf{p}) \mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H}\mathbf{H}\mathbf{p}\mathbf{p}^{\mathrm{T}}\mathbf{H} g_1^2(\mathbf{x}+\mathbf{H}\mathbf{p}) g_2^2(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{p}) d\mathbf{p} \\ &= o(\mathbf{1}_{d\times d}).\end{aligned}$$

Proposition A.1 *Under assumption* (A3)*, there exists a constant* A*, such that,*

$$n^2|\mathbf{H}^2| \iiint_{\|\mathbf{p}\|\leq 1, \|\mathbf{q}\|\leq 1, \|\mathbf{r}\|\leq 1} |\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r})]| d\mathbf{p} d\mathbf{q} d\mathbf{r} \leq A.$$

Proof of Proposition A.1 The proof can be found in Liu (2001, Proposition 2.2).

Proposition A.2 *Under assumptions* (A1)–(A3), (H2), (K1) and (K2),

$$\lim_{n\to\infty} n|\mathbf{H}| \int K(\mathbf{p})K(\mathbf{q})\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q} = R(K)\rho_c.$$

Proof of Proposition A.2 The proof can be found in Liu (2001, Proposition 2.3).

Lemma A.2 *Let*

$$\begin{aligned}
\tilde{s}_{1,n}(\mathbf{x}) &= \frac{1}{n^2}\sum_{i\neq j}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\, g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j),\\
\tilde{s}_{2,n}(\mathbf{x}) &= \frac{1}{n^2}\sum_{i\neq j}^{n} |K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\, g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)|,
\end{aligned}$$

where g_1 and g_2 are bounded functions at $\mathbf{x}$, under assumptions (A1)–(A3), (H2), (K1) and (K2)*, if $\mathbf{x}$ is an interior point of the support of f, then,*

$$\tilde{s}_{1,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}R(K)f^2(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x})\rho_c + o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right), \tag{A.7}$$

$$\tilde{s}_{2,n}(\mathbf{x}) = \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right). \tag{A.8}$$

Proof of Lemma A.2 First, we will prove expression (A.7). Notice that

$$\begin{aligned}
\mathbb{E}[n|\mathbf{H}|\tilde{s}_{1,n}(\mathbf{x})] &= \mathbb{E}\left[\frac{|\mathbf{H}|}{n}\sum_{i\neq j}^{n} K_{\mathbf{H}}(\mathbf{X}_i-\mathbf{x})\, K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})\, g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)\right]\\
&= (n-1)|\mathbf{H}| \iint K(\mathbf{p})\, K(\mathbf{q})\, \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] f(\mathbf{x}+\mathbf{H}\mathbf{p}) f(\mathbf{x}+\mathbf{H}\mathbf{q})\\
&\quad \cdot g_1(\mathbf{x}+\mathbf{H}\mathbf{p})g_2(\mathbf{x}+\mathbf{H}\mathbf{q})d\mathbf{p}d\mathbf{q}\\
&= \frac{n-1}{n} f^2(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x}) \left\{ n|\mathbf{H}| \iint K(\mathbf{p})\, K(\mathbf{q})\, \rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q} \right\}\\
&\quad \cdot [1+o(1)].
\end{aligned}$$

Using Proposition A.2, it follows that

$$\mathbb{E}[n|\mathbf{H}|\tilde{s}_{1,n}(\mathbf{x})] = R(K)f^2(\mathbf{x})g_1(\mathbf{x})g_2(\mathbf{x})\rho_c + o(1).$$

For the variance, note that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[n|\mathbf{H}|\tilde{s}_{1,n}(\mathbf{x})] &= \mathbb{V}\mathrm{ar}\left[\frac{|\mathbf{H}|}{n}\sum_{i\neq j}^{n} K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)\right] \\ &= 4n^{-2}|\mathbf{H}|^2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{k=1}^{n-1}\sum_{l=k+1}^{n}\mathbb{C}\mathrm{ov}[\chi_{ij}(\mathbf{x}),\chi_{kl}(\mathbf{x})], \qquad (A.9)\end{aligned}$$

where

$$\chi_{ij}(\mathbf{x}) = K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{x}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j).$$

Now, consider the value of $\mathbb{C}\mathrm{ov}[\chi_{ij}(\mathbf{x}),\chi_{kl}(\mathbf{x})]$ according to the following three exclusive cases: when $i=k$ and $j=l$, when $i=k$ and $j\neq l$, and, finally, when i,j,k,l are all distinct.

First, when $i=k$ and $j=l$, the total number of such terms is $n(n-1)/2$. In this case, one gets

$$\begin{aligned}\mathbb{C}\mathrm{ov}[\chi_{ij}(\mathbf{x}),\chi_{ij}(\mathbf{x})] &\leq \mathbb{E}\left[K_{\mathbf{H}}^2\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}^2\left(\mathbf{X}_j-\mathbf{x}\right)g_1^2(\mathbf{X}_i)g_2^2(\mathbf{X}_j)\rho_n^2(\mathbf{X}_i-\mathbf{X}_j)\right] \\ &= |\mathbf{H}|^{-2}f^2(\mathbf{x})g_1^2(\mathbf{x})g_2^2(\mathbf{x})\iint K^2\left(\mathbf{p}\right)K^2\left(\mathbf{q}\right)\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}d\mathbf{q} \\ &\quad \cdot[1+o(1)] \\ &\leq \frac{K_M^4 f_M^2 g_{1,M}^2 g_{2,M}^2}{n|\mathbf{H}|^3}\int_{\|\mathbf{q}\|\leq 1}\{n|\mathbf{H}|\int_{\|\mathbf{p}\|\leq 1}\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p}\}d\mathbf{q} \\ &\quad \cdot[1+o(1)],\end{aligned}$$

where f_M, $g_{1,M}$, $g_{2,M}$ and K_M denote the upper bounds of f, g_1, g_2 and K, respectively.

Since

$$n|\mathbf{H}|\int_{\|\mathbf{p}\|\leq 1}\rho_n^2[\mathbf{H}(\mathbf{p}-\mathbf{q})]d\mathbf{p} \leq n\int|\rho_n(\mathbf{t})|d\mathbf{t} \leq A_1,$$

then

$$\begin{aligned}\mathbb{C}\text{ov}[\chi_{ij}(\mathbf{x}),\chi_{ij}(\mathbf{x})] &\leq \frac{K_M^4 f_M^2 g_{1,M}^2 g_{2,M}^2 A_1}{n|\mathbf{H}|^3} \times \text{Volume of the ball } \{\mathbf{u}:\|\mathbf{u}\|\leq 1\} \\ &= \frac{A_2}{n|\mathbf{H}|^3}. \end{aligned} \tag{A.10}$$

Second, when $i=k$ and $j\neq l$ in the expression (A.9). Notice that in this case, the total number of such terms can be bounded by n^3. It follows that

$$\begin{aligned}\mathbb{C}\text{ov}[\chi_{ij}(\mathbf{x}),\chi_{il}(\mathbf{x})] &= \mathbb{E}[\chi_{ij}(\mathbf{x})\chi_{il}(\mathbf{x})] - \mathbb{E}[\chi_{ij}(\mathbf{x})]\mathbb{E}[\chi_{il}(\mathbf{x})] \\ &= \mathbb{E}[\chi_{ij}(\mathbf{x})\chi_{il}(\mathbf{x})] - [\mathbb{E}[\chi_{ij}(\mathbf{x})]]^2 \\ &\leq \mathbb{E}[\chi_{ij}(\mathbf{x})\chi_{il}(\mathbf{x})] \\ &= \mathbb{E}\Big[K_{\mathbf{H}}^2(\mathbf{X}_i-\mathbf{x})\,K_{\mathbf{H}}(\mathbf{X}_j-\mathbf{x})K_{\mathbf{H}}(\mathbf{X}_l-\mathbf{x})\,g_1^2(\mathbf{X}_i)g_2(\mathbf{X}_j)g_2(\mathbf{X}_l) \\ &\quad \cdot\rho_n(\mathbf{X}_i-\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_l)\Big] \\ &= |\mathbf{H}|^{-1}f^3(\mathbf{x})g_1^2(\mathbf{x})g_2^2(\mathbf{x})\iiint K^2(\mathbf{p})\,K(\mathbf{q})\,K(\mathbf{r})\,\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})] \\ &\quad \cdot\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r})]d\mathbf{p}d\mathbf{q}d\mathbf{r}\cdot[1+o(1)] \\ &\leq \frac{K_M^4 f_M^3 g_{1,M}^2 g_{2,M}^2}{n^2|\mathbf{H}|^3} \\ &\quad \cdot\{n^2|\mathbf{H}^2|\iiint_{\|\mathbf{p}\|\leq 1,\|\mathbf{q}\|\leq 1,\|\mathbf{r}\|\leq 1}|\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{q})]\rho_n[\mathbf{H}(\mathbf{p}-\mathbf{r})]|d\mathbf{p}d\mathbf{q}d\mathbf{r}\} \\ &\quad \cdot[1+o(1)]. \end{aligned}$$

Using Proposition A.1, it follows that

$$\begin{aligned}\mathbb{C}\text{ov}[\chi_{ij}(\mathbf{x}),\chi_{il}(\mathbf{x})] &\leq \frac{K_M^4 f_M^3 g_{1,M}^2 g_{2,M}^2 A}{n^2|\mathbf{H}|^3} \\ &= \frac{A_3}{n|\mathbf{H}|^3}. \end{aligned} \tag{A.11}$$

Finally, when i,j,k,l are all distinct in (A.9), given that ϕ_{ij} and ϕ_{kl} are independent,

$$\mathbb{C}\text{ov}[\chi_{ij}(\mathbf{x}),\chi_{kl}(\mathbf{x})] = 0. \tag{A.12}$$

Then, considering (A.9), (A.10), (A.11) and (A.12), it follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}[n|\mathbf{H}|\tilde{s}_{1,n}(\mathbf{x})] &= 4n^{-2}|\mathbf{H}|^2\left(\frac{n^2-n}{2}\frac{A_2}{n|\mathbf{H}|^3}+n^3\frac{A_3}{n^2|\mathbf{H}|^3}\right)\\ &= o(1).\end{aligned}$$

Using similar arguments, expression (A.8) can be directly proved.

Lemma A.3 *Let*

$$\begin{aligned}\tilde{s}_{3,n}(\mathbf{x}) &= \frac{1}{n^2}\sum_{i\neq j}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{x}),\\ \tilde{s}_{4,n}(\mathbf{x}) &= \frac{1}{n^2}\sum_{i\neq j}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{x})\\ &\quad\cdot(\mathbf{X}_j-\mathbf{x})^{\mathrm{T}},\end{aligned}$$

where g_1 and g_2 are bounded functions at $\mathbf{x}$, under assumptions (A1)–(A3), (H2), (K1) and (K2)*, if $\mathbf{x}$ is an interior point of the support of f, then,*

$$\tilde{s}_{3,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}o_{\mathbb{P}}(\mathbf{1}_d), \tag{A.13}$$

$$\tilde{s}_{4,n}(\mathbf{x}) = \frac{1}{n|\mathbf{H}|}o_{\mathbb{P}}(\mathbf{1}_{d\times d}). \tag{A.14}$$

Proof of Lemma A.3 First, expression (A.13) is proved. Notice that, expression (A.13) is equivalent to

$$\mathbf{v}^{\mathrm{T}}\tilde{s}_{3,n}(\mathbf{x}) = o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right),$$

for any $\mathbf{v}\in\mathbb{R}^d$.

First,

$$\begin{aligned}\mathbf{v}^{\mathrm{T}}\tilde{s}_{3,n}(\mathbf{x}) &= \frac{1}{n^2}\sum_{i\neq j}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)\mathbf{v}^{\mathrm{T}}\left(\mathbf{X}_i-\mathbf{x}\right)\\ &\quad\cdot\mathbb{I}_{\{\|\mathbf{H}^{-1}(\mathbf{X}_i-\mathbf{x})\|\leq 1\}}.\end{aligned}$$

Note that

$$|\mathbf{v}^{\mathrm{T}}\tilde{s}_{3,n}(\mathbf{x})| \leq \|\mathbf{v}\|\|\mathbf{H}\|\frac{1}{n^2}\sum_{i\neq j}|K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)|,$$

where $\|\cdot\|$ denotes the matrix norm.

Using (A.8) of Lemma A.2, it can be obtained that

$$\begin{aligned}|\mathbf{v}^{\mathrm{T}}\tilde{s}_{3,n}(\mathbf{x})| &= \|\mathbf{v}\|\|\mathbf{H}\|\mathcal{O}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right)\\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{\|\mathbf{H}\|}{n|\mathbf{H}|}\right)\\ &= o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

Now, expression (A.14) is proved. Note that, (A.14) is equivalent to

$$\mathbf{v}_1^{\mathrm{T}}\tilde{s}_{4,n}(\mathbf{x})\mathbf{v}_2 = o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right),$$

for any $\mathbf{v}_1,\mathbf{v}_2\in\mathbb{R}^d$.

Notice that

$$\begin{aligned}\mathbf{v}_1^{\mathrm{T}}\tilde{s}_{4,n}(\mathbf{x})\mathbf{v}_2 &= \frac{1}{n^2}\sum_{i\neq j}K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)\\ &\quad\cdot\mathbf{v}_1^{\mathrm{T}}\left(\mathbf{X}_i-\mathbf{x}\right)\left(\mathbf{X}_j-\mathbf{x}\right)^{\mathrm{T}}\mathbf{v}_2\mathbb{I}_{\{\|\mathbf{H}^{-1}(\mathbf{X}_i-\mathbf{x})\|\leq 1\}}\end{aligned}$$

and, therefore,

$$|\mathbf{v}_1^{\mathrm{T}}\tilde{s}_{4,n}(\mathbf{x})\mathbf{v}_2| \leq \|\mathbf{v}\|\|\mathbf{H}\|^2\frac{1}{n^2}\sum_{i\neq j}|K_{\mathbf{H}}\left(\mathbf{X}_i-\mathbf{x}\right)K_{\mathbf{H}}\left(\mathbf{X}_j-\mathbf{t}\right)g_1(\mathbf{X}_i)g_2(\mathbf{X}_j)\rho_n(\mathbf{X}_i-\mathbf{X}_j)|.$$

Using (A.8) of Lemma A.2, it can be obtained that

$$\begin{aligned}|\mathbf{v}_1^{\mathrm{T}}\tilde{s}_{4,n}(\mathbf{x})\mathbf{v}_2| &= \|\mathbf{v}\|\|\mathbf{H}\|^2\mathcal{O}_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right)\\ &= o_{\mathbb{P}}\left(\frac{1}{n|\mathbf{H}|}\right).\end{aligned}$$

A.2 Approximation of a summation by an integral

For a fixed design, an approximation of a summation by an integral is provided. More specifically, the approximation of a summation by an integral involving a kernel function is given in Lemma A.4. For other approximations, an analogous procedure to that used in the proof of Lemma A.4 can be used. These approximations are employed in the proof of Theorem 2.2.

Lemma A.4 *For any* $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$*, under assumption* (K1)*, it follows that*

$$\sum_{i=1}^{n} n^{-1} K_{\mathbf{H}}\left(\mathbf{x}_i - \mathbf{x}\right) = \int K(\mathbf{p}) d\mathbf{p} + o(1).$$

Proof of Lemma A.4 Let $\phi(\mathbf{x}_i) = K[\mathbf{H}^{-1}(\mathbf{x}_i - \mathbf{x})]$ being $\mathbf{x} = (x_1, x_2)$ and $\mathbf{x}_i = (x_{i1}, x_{i2})$. For $i = 1, \ldots, n$, consider the multivariate Ostrowski's inequality (Anastassiou, 1997) provided in Theorem A.1 in Section A.3, where $a_j = a_{ij} = x_{ij} - l_{ij}/2$, $b_j = b_{ij} = x_{ij} + l_{ij}/2$, with $l_{ij} = b_{ij} - a_{ij} = K_j/\sqrt{n} + o\left(K_j/\sqrt{n}\right)$, being $K_j = \max_i x_{ij}$, for $j = 1, 2$. With this choice, it follows that

$$\begin{aligned} &\left| \phi(x_{i1}, x_{i2}) - \frac{1}{l_{i1} l_{i2}} \int_{a_{i1}}^{b_{i1}} \int_{a_{i2}}^{b_{i2}} \phi(u_1, u_2) du_2 du_1 \right| \\ &\qquad \leq \frac{1}{4} \left(l_{i1} \left\| \frac{\partial \phi}{\partial u_1} \right\|_\infty + l_{i2} \left\| \frac{\partial \phi}{\partial u_2} \right\|_\infty \right). \end{aligned} \tag{A.15}$$

For the sake of simplicity, a regular design in the unit square is assumed, and consequently, $K_1 = K_2 = 1$ and $l_1 = l_2 = 1/\sqrt{n}$. Using (A.15) and denoting $\xi_i = \frac{1}{4}\left[l_1 \left\| \frac{\partial \phi}{\partial u_1} \right\|_\infty + l_2 \left\| \frac{\partial \phi}{\partial u_2} \right\|_\infty \right]$, it follows that

$$\begin{aligned} \sum_{i=1}^{n} \phi(x_{i1}, x_{i2}) &= \sum_{i=1}^{n} \frac{1}{l_{i1} l_{i2}} \int_{a_{i1}}^{b_{i1}} \int_{a_{i2}}^{b_{i2}} \phi(u_1, u_2) du_2 du_1 + \sum_{i=1}^{n} \xi_i. \\ &= n \int_{\mathcal{D}_1} \int_{\mathcal{D}_2} \phi(u_1, u_2) du_2 du_1 + \sum_{i=1}^{n} \xi_i \\ &= n \int \phi(\mathbf{u}) d\mathbf{u} + \sum_{i=1}^{n} \xi_i. \end{aligned} \tag{A.16}$$

Assume that $\mathbf{H} = \text{diag}(h_1, h_2)$. For $\mathbf{u} = (u_1, u_2)$, it is obtained that

$$\mathbf{H}^{-1}(\mathbf{u} - \mathbf{x}) = \begin{pmatrix} h_1^{-1}(u_1 - x_1) \\ h_2^{-1}(u_2 - x_2) \end{pmatrix}$$

and, therefore,

$$\phi(u_1, u_2) = K \begin{pmatrix} h_1^{-1}(u_1 - x_1) \\ h_2^{-1}(u_2 - x_2) \end{pmatrix}.$$

Denoting $v_1 = h_1^{-1}(u_1 - x_1)$ and $v_2 = h_2^{-1}(u_2 - x_2)$, the partial derivatives of the function $\phi : \mathbb{R}^2 \to \mathbb{R}$ are

$$\begin{aligned} \frac{\partial \phi}{\partial u_1} &= \frac{\partial \phi}{\partial v_1}\frac{\partial v_1}{\partial u_1} + \frac{\partial \phi}{\partial v_2}\frac{\partial v_2}{\partial u_1} = \frac{\partial K}{\partial v_1}\frac{1}{h_1}, \\ \frac{\partial \phi}{\partial u_2} &= \frac{\partial \phi}{\partial v_1}\frac{\partial v_1}{\partial u_2} + \frac{\partial \phi}{\partial v_2}\frac{\partial v_2}{\partial u_2} = \frac{\partial K}{\partial v_2}\frac{1}{h_2}. \end{aligned}$$

Therefore, the error term in (A.16) can be rewritten as

$$\begin{aligned} \sum_{i=1}^{n} \xi_i &= \frac{1}{4}\sum_{i=1}^{n}\left[l_{i1} \left\| \frac{\partial \phi}{\partial u_1} \right\|_\infty + l_{i2} \left\| \frac{\partial \phi}{\partial u_2} \right\|_\infty \right] \\ &= \frac{1}{4}\sum_{i=1}^{n}\left[l_{i1} \left\| \frac{\partial K}{\partial v_1}\frac{1}{h_1} \right\|_\infty + l_{i2} \left\| \frac{\partial K}{\partial v_2}\frac{1}{h_2} \right\|_\infty \right]. \end{aligned}$$

If $l_{ij} = n^{-1/2}$, for $j = 1, 2$, it follows that

$$\begin{aligned} \sum_{i=1}^{n} \xi_i &= \frac{1}{4} n \left[\mathcal{O}\left(\frac{1}{h_1}\frac{1}{\sqrt{n}} \right) + \mathcal{O}\left(\frac{1}{h_2}\frac{1}{\sqrt{n}} \right) \right] \\ &= \mathcal{O}\left(\frac{\sqrt{n}}{|\mathbf{H}|^{1/2}} \right). \end{aligned} \tag{A.17}$$

Using (A.16) and (A.17), it can be obtained that

$$\begin{aligned} \sum_{i=1}^{n} \phi(x_{i1}, x_{i2}) &= n \int \phi(\mathbf{u}) d\mathbf{u} + o(n) \\ &= n \int \phi(\mathbf{u}) d\mathbf{u} \cdot [1 + o(1)]. \end{aligned}$$

Consequently,

$$\begin{aligned}\sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x}_i - \mathbf{x}) &= n|\mathbf{H}|^{-1} \int K[\mathbf{H}^{-1}(\mathbf{u}-\mathbf{x})]d\mathbf{u} \cdot [1 + o(1)] \\ &= n \int K(\mathbf{p})d\mathbf{p} \cdot [1 + o(1)] \\ &= n \cdot [1 + o(1)].\end{aligned}$$

A.3 Auxiliary results

This sections contains other results which are employed throughout this thesis.

Proposition A.3 *Let* $\mathcal{S} = \{\mathbf{B} : \mathbf{B}$ *is a* $d \times d$ *symmetric, positive definite matrix*$\}$. *Assuming that* $\mathbf{C}$ *is a* $d \times d$ *symmetric matrix (either positive definite or negative definite), and* c_1, c_2 *are positive numbers, then the solution to the optimization problem*

$$\min_{\mathbf{B} \in \mathcal{S}_m} L(\mathbf{B}) = \left\{ c_1 \mathrm{tr}^2(\mathbf{B}^2\mathbf{C}) + \frac{c_2}{|\mathbf{B}|} \right\}$$

is

$$\mathbf{B} = \left\{ \frac{c_2 |\tilde{\mathbf{C}}|^{1/2}}{4c_1 d} \right\}^{1/(d+4)} \tilde{\mathbf{C}}^{-1/2},$$

with

$$\tilde{\mathbf{C}} = \begin{cases} \mathbf{C} & \text{if} \quad \mathbf{C} \text{ is positive definite,} \\ -\mathbf{C} & \text{if} \quad \mathbf{C} \text{ is negative definite.} \end{cases}$$

Proof of Proposition A.3 The proof of this result can be found in Liu (2001, Proposition 2.6).

Theorem A.1 (Multivariate Ostrowski's inequality) *Let* f *be a continuously differentiable function in* $\prod_{j=1}^{d}[a_j, b_j]$, *where* $a_j < b_j$, *with* $a_j, b_j \in \mathbb{R}$, $j = 1, \cdots, d$, *and let* $\mathbf{x}_0 = (x_{01}, \cdots, x_{0d}) \in \prod_{j=1}^{d}[a_j, b_j]$ *be fixed. Then,*

$$\begin{aligned}&\left| f(\mathbf{x}_0) - \frac{1}{\prod_{j=1}^{d}(b_j - a_j)} \int_{a_1}^{b_1} \cdots \int_{a_d}^{b_d} f(z_1, \ldots, z_d) dz_1 \ldots dz_d \right| \\ &\leq \sum_{j=1}^{d} \left(\frac{(x_{0j} - a_j)^2 + (b_j - x_{0j})^2}{2(b_j - a_j)} \right) \left\| \frac{\partial f}{\partial z_j} \right\|_{\infty}.\end{aligned}$$

Proof of Theorem A.1 The proof of this result can be found in Anastassiou (1997).

Resumen en castellano

Para analizar la dependencia de una variable de interés (variable respuesta o variable dependiente) con respecto a otra(s) variable(s) (covariables, predictores, variables explicativas o variables independientes), se pueden utilizar los modelos de regresión. En general, en esta tesis se consideran modelos de regresión múltiples (respuesta univariante y predictor multivariante) que presentan un cierto tipo de estructura de dependencia espacial. En esta situación, para poder realizar inferencias fiables, tanto la función de regresión como la estructura de dependencia, que suelen ser desconocidas, deben especificarse adecuadamente.

La estimación de la función de regresión puede ser abordada empleando métodos no paramétricos, obteniéndose estimadores flexibles y evitando problemas de mala especificación. Alternativamente, también podrían utilizarse métodos paramétricos. Los procedimientos paramétricos deberían ser empleados si la función de regresión pertenece a la familia paramétrica asumida. Sin embargo, una mala especificación de esa familia puede llevar a conclusiones equivocadas. Los problemas de especificación incorrecta de la función de regresión se pueden evitar aplicando un contraste de bondad de ajuste para el modelo seleccionado. Para datos que presentan algún tipo de complejidad, por ejemplo, datos circulares, los procedimientos clásicos utilizados en regresión para datos Euclídeos no pueden emplearse directamente, sino que deben adaptarse y modificarse convenientemente.

El objetivo de esta tesis es doble. Primero, se analizan algunos problemas de inferencia para modelos de regresión con respuesta y covariables Euclídeas en presencia de dependencia espacial. Concretamente, se presenta y se analiza un procedimiento para contrastar si la función de regresión pertenece a una familia paramétrica, en presencia de correlación espacial. El segundo objetivo es diseñar y estudiar nuevos

procedimientos para abordar problemas de estimación y contrastes de la función regresión (circular) para modelos con respuesta circular y covariable con valores en $\mathbb{R}^d$. En este contexto, se presentan y estudian propuestas no paramétricas para estimar la función de regresión, bajo el supuesto de independencia y también para errores espacialmente correlados. Además, en estos dos contextos, se presentan contrastes de bondad de ajuste para determinar si la función de regresión circular pertenece a una familia paramétrica.

A continuación se incluye un resumen con los distintos temas tratados a lo largo de la tesis, así como las principales aportaciones realizadas en cada capítulo de la misma. Las demostraciones de los distintos resultados teóricos se pueden ver al final de cada capítulo.

Capítulo 1: Introducción.

En este capítulo se describen algunos métodos existentes para resolver ciertos problemas de inferencia estadística (estimación y contrastes de bondad de ajuste) que involucran datos Euclídeos y circulares.

En primer lugar, se proporciona una revisión acerca de diferentes procedimientos (paramétricos y no paramétricos) para estimar la función de regresión en modelos con respuesta y covariables Euclídeas, tanto para datos independientes como para datos espacialmente correlados.

Sea $\{(\mathbf{X}_i, Z_i)\}_{i=1}^n$ una muestra aleatoria de $(\mathbf{X}, Z)$, donde Z es una variable respuesta escalar que depende de d covariables (deterministas o aleatorias) $\mathbf{X}$, con soporte $\mathcal{D} \subset \mathbb{R}^d$. Supongamos el siguiente modelo de regresión:

$$Z_i = m(\mathbf{X}_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

donde m es la función de regresión o tendencia y $\varepsilon_i, i = 1, \ldots, n$, son variables aleatorias generadas a partir de un proceso de media cero y estacionario de segundo orden con covariograma o función de covarianza $\mathbb{C}\mathrm{ov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \rho_n(\mathbf{X}_i - \mathbf{X}_j), i, j = 1, \ldots, n$, donde σ^2 es la varianza de los errores y ρ_n es una función de correlación estacionaria y continua tal que $\rho_n(\mathbf{0}) = 1$, $\rho_n(\mathbf{x}) = \rho_n(-\mathbf{x})$ y $|\rho_n(\mathbf{x})| \leq 1$, $\forall \mathbf{x} \in \mathcal{D}$.

Suponiendo que la función de regresión pertenece a una familia paramétrica,

$m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}$, donde $\boldsymbol{\mathcal{B}} \subset \mathbb{R}^q$ es un conjunto compacto, y q es la dimensión del espacio de parámetros $\boldsymbol{\mathcal{B}}$, un estimador paramétrico de m podría obtenerse empleando métodos basados en mínimos cuadrados o máxima verosimilitud. En un contexto de errores espacialmente correlados, se puede considerar el estimador obtenido a partir de mínimos cuadrados iterados, que tiene en cuenta la estructura de dependencia espacial subyacente (Newman y Jacobson, 1984). Alternativamente, sin suponer una forma paramétrica para la función de regresión, esta podría estimarse no paramétricamente, por ejemplo, utilizando métodos tipo núcleo (*kernel*). En esta tesis, nos centraremos principalmente en los estimadores de Nadaraya–Watson (o estimador polinómico local de grado cero) y lineal local. En el caso multivariante, estos estimadores dependen de una función kernel multivariante y de una matriz de suavizado o matriz ventana. La elección de esta matriz es crucial en el procedimiento de estimación. Para datos independientes, las propiedades asintóticas del estimador de Nadaraya–Watson fueron calculadas por Härdle y Müller (2012), mientras que las del estimador lineal local fueron estudiadas por Ruppert y Wand (1994). En un contexto de errores espacialmente correlados, el sesgo y la varianza asintóticas de estos estimadores fueron calculadas por Liu (2001).

Como se mencionó previamente, los métodos paramétricos pueden ser preferibles si la función de regresión pertenece a la familia paramétrica asumida. Sin embargo, estos procedimientos pueden proporcionar resultados erróneos si el modelo no se especifica correctamente. Para determinar si una familia paramétrica es adecuada para modelar la función de regresión, se pueden aplicar contrastes de bondad de ajuste sobre el modelo seleccionado. El contraste de hipótesis que se formula es el siguiente:

$$H_0 : m \in \mathcal{M}_{\boldsymbol{\beta}} = \{m_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}, \qquad \text{vs.} \qquad H_a : m \notin \mathcal{M}_{\boldsymbol{\beta}}. \tag{2}$$

Este problema ha sido estudiado en diferentes contextos a lo largo de la literatura, comparando un estimador no paramétrico y un estimador paramétrico de la función de regresión bajo la hipótesis nula (Härdle y Mammen, 1993; Alcalá et al., 1999). Sin embargo, este no había sido considerado en el caso de datos con dependencia espacial. En el Capítulo 2 de esta tesis se aborda este problema para modelos de regresión con respuesta y covariables Euclídeas en presencia de dependencia espacial.

En segundo lugar, en el Capítulo 1 también se introduce el modelo de regresión para datos circulares que se supone en esta memoria, bajo la hipótesis de independencia (Capítulos 3 y 5) y para datos espacialmente correlados (Capítulos 4 y 5). Más concretamente, si $\{(\mathbf{X}_i, \Theta_i)\}_{i=1}^n$ es una muestra de $(\mathbf{X}, \Theta)$, donde Θ es una variable aleatoria circular que toma valores en $\mathbb{T} = [0, 2\pi)$, y $\mathbf{X}$ es una variable aleatoria definida en $\mathcal{D} \subseteq \mathbb{R}^d$, se considera el modelo:

$$\Theta_i = [m^c(\mathbf{X}_i) + \varepsilon_i](\mathtt{mod}\, 2\pi), \quad i = 1, \ldots, n, \tag{3}$$

donde m^c es una función de regresión circular y $\varepsilon_i, i = 1, \ldots, n$, son variables aleatorias circulares con dirección media cero y concentración finita.

En los Capítulos 3, 4 y 5 de esta tesis, considerando el modelo de regresión (3), se abordan ciertos problemas de inferencia estadística. Concretamente, en los Capítulos 3 y 4, se proponen y estudian estimadores no paramétricos de la función de regresión circular m^c, para datos independientes y para datos espacialmente correlados, respectivamente. Para estos dos contextos, en el Capítulo 5 se presentan contrastes de bondad de ajuste para determinar si la función de regresión circular pertenece a una familia paramétrica.

Capítulo 2: Contrastes de bondad de ajuste para funciones de regresión paramétricas con errores espacialmente correlados.

En este capítulo se contrasta si la función de regresión pertenece a una familia paramétrica, considerando modelos de regresión con respuesta y covariables Euclídeas en presencia de dependencia espacial. Es decir, considerando el modelo (1), se resuelve el problema formulado en (2). El estadístico de contraste propuesto compara en términos de una distancia L_2 una versión suavizada de un ajuste paramétrico con un estimador no paramétrico de la función de regresión:

$$T_{n,p} = n|\mathbf{H}|^{1/2} \int_{\mathcal{D}} [\hat{m}_{\mathbf{H}}(\mathbf{x}; p) - \hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)]^2 w(\mathbf{x}) d\mathbf{x}, \quad p = 0, 1, \tag{4}$$

donde w es una función de ponderación que ayuda a mitigar el posible efecto frontera. Los estimadores no paramétricos $\hat{m}_{\mathbf{H}}(\mathbf{x}; p)$ serían el estimador de Nadaraya–Watson

(cuando $p = 0$) y el estimador lineal local (cuando $p = 1$), mientras que $\hat{m}_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)$ serían las correspondientes versiones suavizadas del ajuste paramétrico $m_{\hat{\boldsymbol{\beta}}}$ (obtenido empleando mínimos cuadrados iterados).

El estadístico de contraste dado en (4) depende de la matriz ventana $\mathbf{H}$. La selección de la matriz ventana se ha abordado en la literatura en el contexto de estimación. Sin embargo, este es un problema abierto en el caso de los contrastes de bondad de ajuste. En la práctica, como es habitual en este tipo de contextos, se analiza el comportamiento del estadístico de contraste en una rejilla de ventanas $\mathbf{H}$.

En este capítulo se calcula la distribución asintótica del estadístico de contraste propuesto (para diseño fijo y aleatorio) bajo la hipótesis nula y también bajo alternativas de Pitman, generalizando los resultados obtenidos por Härdle y Mammen (1993) y Alcalá et al. (1999) en el caso univariante y con errores independientes. Dado que la convergencia a la distribución asintótica es lenta y, además, esta depende de elementos desconocidos, no suele emplearse en la práctica. Para calibrar el contraste, se proponen tres métodos bootstrap: un bootstrap paramétrico (PB), un bootstrap no paramétrico (NPB) y un bootstrap no paramétrico corregido (CNPB). En PB se utilizan los residuos obtenidos a partir del ajuste paramétrico y, a partir de estos, se estima paramétricamente la estructura de dependencia. En NPC, para aumentar la potencia del contraste, se consideran los residuos a partir del ajuste no paramétrico (González-Manteiga y Cao, 1993). En cuanto a CNPB, este procedimiento es una modificación de NPB que corrige el sesgo obtenido debido a la estimación no paramétrica del variograma.

El comportamiento empírico del contraste se comprueba a partir de un extenso estudio de simulación. En este estudio se muestra que bajo la hipótesis nula, tanto PB como CNPB funcionan correctamente, pues las proporciones de rechazos son similares al nivel de significación considerado. Además, cuando la estructura de dependencia es más fuerte (el valor del rango de dependencia de las observaciones crece), la matriz ventana $\mathbf{H}$ que proporciona una calibración efectiva debe ser más grande. Bajo hipótesis alternativas, el único procedimiento que funciona correctamente es CNPB, ya que el comportamiento de PB es bastante pobre. Cabe destacar que, aunque NPB proporciona una potencia alta, este procedimiento no proporciona un calibrado efectivo, ya que las proporciones de rechazos bajo la hipótesis nula son

excesivamente altas. Una vez comprobado el correcto funcionamiento de CNPB en el estudio de simulación, se aplica el contraste propuesto al ejemplo del acuífero de Wolfcamp, un conjunto de datos clásico en la literatura geoestadística.

Las contribuciones de este capítulo se pueden ver en Meilán-Vila et al. (2020e) y Meilán-Vila et al. (2020b). En Meilán-Vila et al. (2020e), se considera un diseño aleatorio y el estimador lineal local para el ajuste no paramétrico. Por otro lado, en Meilán-Vila et al. (2020b), bajo diseño fijo y considerando el estimador de Nadaraya-Watson, se proporciona un detallado análisis computacional del comportamiento del contraste usando diferentes algoritmos bootstrap.

Capítulo 3: Estimación no paramétrica de la función de regresión con respuesta circular y covariable con valores en $\mathbb{R}^d$.

El objetivo de este capítulo es estimar la función de regresión para modelos de regresión con respuesta circular y predictor Euclídeo multidimensional.

Considerando el modelo (3), la función de regresión circular puede definirse como el minimizador de la función de riesgo $\mathbb{E}\{1-\cos[\Theta-m^c(\mathbf{X})] \mid \mathbf{X}=\mathbf{x}\}$. Puede comprobarse que el minimizador de esa función de riesgo es $m^c(\mathbf{x}) = \operatorname{atan2}[m_1(\mathbf{x}), m_2(\mathbf{x})]$, donde $m_1(\mathbf{x}) = \mathbb{E}[\sin(\Theta) \mid \mathbf{X}=\mathbf{x}]$ y $m_2(\mathbf{x}) = \mathbb{E}[\cos(\Theta) \mid \mathbf{X}=\mathbf{x}]$. La función $\operatorname{atan2}(y,x)$ devuelve el ángulo entre el eje x y el vector desde el origen hasta (x,y). Substituyendo m_1 y m_2 por dos estimadores apropiados, se puede obtener un estimador para la función de regresión circular m^c. Por lo tanto, considerando estimadores polinómicos locales para m_1 y para m_2, se tiene que el estimador de tipo polinómico local de la función de regresión es:

$$\hat{m}^c_{\mathbf{H}}(\mathbf{x};p) = \operatorname{atan2}[\hat{m}_{1,\mathbf{H}}(\mathbf{x};p), \hat{m}_{2,\mathbf{H}}(\mathbf{x};p)], \tag{5}$$

donde para cualquier entero $p \geq 0$, $\hat{m}_{1,\mathbf{H}}(\mathbf{x};p)$ y $\hat{m}_{2,\mathbf{H}}(\mathbf{x};p)$ denotan los estimadores polinómico locales de grado p (con matriz ventana $\mathbf{H}$) de $m_1(\mathbf{x})$ y $m_2(\mathbf{x})$, respectivamente. Cuando el grado del polinomio es cero, se tendría el estimador de tipo Nadaraya–Watson, mientras que cuando es uno, se obtendría el estimador de tipo lineal local.

En este capítulo, se calculan las propiedades asintóticas del estimador (5) cuando

$p = 0$ y $p = 1$. El sesgo y la varianza de estos estimadores tienen una estructura similar a los obtenidos para el estimador de Nadaraya–Watson (en el caso de $p = 0$) o el lineal local (si $p = 1$) cuando la respuesta del modelo de regresión es Euclídea. El sesgo de ambos estimadores es diferente (como ocurre en el caso Euclídeo), ya que en el caso del estimador de tipo lineal local no depende de la densidad del diseño. Por otra parte, el término principal de la varianza es el mismo para $p = 0$ y $p = 1$. Estos resultados generalizan los que obtuvieron Di Marzio et al. (2013) en el caso de un único predictor. Para $p > 1$, procediendo como en Ruppert y Wand (1994), se estudian los estimadores $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)$ cuando $d = 1$.

Para comprobar el funcionamiento práctico del estimador (5), se realiza un completo estudio de simulación considerando diferentes escenarios. La matriz ventana se elige empleando el criterio de validación cruzada adaptada a un contexto circular, seleccionando la matriz $\mathbf{H}_{\mathrm{CV}^c}$ que minimiza la función:

$$\mathrm{CV}^c(\mathbf{H}) = \sum_{i=1}^{n} \left\{1 - \cos\left[\Theta_i - \hat{m}^c_{\mathbf{H},-i}(\mathbf{X}_i; p)\right]\right\},$$

donde $\hat{m}^c_{\mathbf{H},-i}(\mathbf{X}_i; p)$ es el estimador de tipo Nadaraya–Watson ($p = 0$) o lineal local ($p = 1$), calculado usando todas las observaciones excepto $(\mathbf{X}_i, \Theta_i)$ y evaluado en $\mathbf{X}_i$. Para cada escenario, se calcula la media aritmética del error cuadrático promedio circular (CASE), definido como (Kim y SenGupta, 2017):

$$\mathrm{CASE}[\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)] = \frac{1}{n} \sum_{i=1}^{n} \left\{1 - \cos\left[m^c(\mathbf{X}_i) - \hat{m}^c_{\mathbf{H}}(\mathbf{X}_i; p)\right]\right\}, \tag{6}$$

para $p = 0$ (Nadaraya–Watson) y $p = 1$ (lineal local). A efectos comparativos también se calcula la matriz óptima $\mathbf{H}_{\mathrm{CASE}}$ que minimiza el CASE. Aunque ambos estimadores presentan un comportamiento similar, los resultados obtenidos para el estimador de tipo lineal local son ligeramente mejores. El procedimiento de estimación también se aplica a un conjunto de datos que contiene las direcciones de escape de dos especies de pulgas (Scapini et al., 2002).

Las contribuciones de este capítulo se pueden ver en Meilán-Vila et al. (2020d).

Capítulo 4: Estimación no paramétrica de la función de regresión con respuesta circular y errores espacialmente correlados.

En este capítulo se aborda el problema de estimar la función de regresión para modelos con respuesta circular y covariables Euclídeas en presencia de dependencia espacial.

Considerando el modelo (3) con errores espacialmente correlados, la función de regresión se podría estimar empleando el estimador (5) propuesto para datos independientes, seleccionando la matriz ventana de manera adecuada. En este contexto, se calculan las propiedades asintóticas del estimador (5) para $p = 0$ y $p = 1$, obteniéndose que el sesgo del estimador es el mismo que en el caso de datos independientes. Sin embargo, como es esperado, la varianza depende del modelo de dependencia de los errores, y es la misma cuando $p = 0$ y $p = 1$. Además, la varianza del estimador tiene la misma estructura que la que se obtuvo para el estimador de Nadaraya–Watson o el lineal local de la función de regresión en modelos con respuesta Euclídea y errores espacialmente correlados (Liu, 2001).

El comportamiento empírico del estimador se muestra en un amplio estudio de simulación. En este punto, la dificultad surge en la generación de un proceso circular en el que se pueda controlar la estructura de dependencia espacial. En el estudio de simulación realizado, los errores del modelo se generan a partir de procesos normales enrollados y proyectados. Se puede comprobar que ambos procesos transfieren correctamente la estructura de dependencia del proceso lineal al círculo. Para la selección de la matriz ventana, dado que los métodos de validación cruzada no deben utilizarse para datos dependientes, se diseña una versión modificada (MCV^c), que tiene en cuenta la estructura de dependencia subyacente. El criterio propuesto selecciona la matriz ventana $\mathbf{H}$ que minimiza la función:

$$\mathrm{MCV}^c(\mathbf{H}) = \sum_{i=1}^{n} \left\{1 - \cos\left[\Theta_i - \hat{m}^c_{\mathbf{H},-N_i}(\mathbf{X}_i; p)\right]\right\},$$

donde $\hat{m}^c_{\mathbf{H},-N_i}(\mathbf{X}_i; p)$ es el estimador de tipo Nadaraya–Watson ($p = 0$) o lineal local ($p = 1$), calculado usando todas las observaciones excepto las que se encuentran en un entorno N_i de $\mathbf{X}_i$, y evaluado en $\mathbf{X}_i$. Para aplicar este criterio, se debe seleccionar el entorno N_i. Por simplicidad, se considera el criterio MCV^c cuando $N_i = \{\mathbf{X}_j :$

$\|\mathbf{X}_j - \mathbf{X}_i\| \leq l\}$. Si la estructura de dependencia es fuerte, más observaciones deberán ser omitidas en el procedimiento de selección de la ventana, y por lo tanto, el valor de l deberá ser más grande.

En el estudio de simulación se muestra que el método CV^c no funciona correctamente, pues los promedios del CASE, dado en (6), que se obtienen utilizando esta ventana son muy grandes en comparación con los que se obtienen si se emplea la matriz óptima $\mathbf{H}_{\mathrm{CASE}}$. Se observa un mejor funcionamiento del procedimiento de estimación cuando se usa la versión modificada para seleccionar la matriz ventana. Se puede ver que cuando la estructura de dependencia es más fuerte, el valor de l seleccionado debe ser más grande. Además, se obtiene un comportamiento ligeramente mejor cuando se emplea el estimador de tipo lineal local. El procedimiento de estimación también se aplica al conjunto de datos de direcciones de olas en el mar Adriático (Jona-Lasinio et al., 2019).

Las contribuciones de esta parte se han recopilado en Meilán-Vila et al. (2020a).

Capítulo 5: Contrastes de bondad de ajuste para funciones de regresión paramétricas con respuesta circular y covariable con valores en $\mathbb{R}^d$.

En este capítulo se propone un contraste de bondad de ajuste para determinar si la función de regresión circular en un modelo con respuesta circular y covariable Euclídea multidimensional pertenece a una familia paramétrica predeterminada. Es decir, se resuelve un contraste similar al formulado en (2), pero teniendo en cuenta que, en este caso, la familia paramétrica (circular) es $\mathcal{M}^c_{\boldsymbol{\beta}} = \{m^c_{\boldsymbol{\beta}}, \boldsymbol{\beta} \in \boldsymbol{\mathcal{B}}\}$, donde $m^c_{\boldsymbol{\beta}}$ es un modelo de regresión paramétrico (circular) con vector de parámetros $\boldsymbol{\beta}$. En este caso, el problema que se formula es:

$$H_0 : m^c \in \mathcal{M}^c_{\boldsymbol{\beta}} \qquad \text{vs.} \qquad H_a : m^c \notin \mathcal{M}^c_{\boldsymbol{\beta}}. \tag{7}$$

Se proponen y se analizan empíricamente dos estadísticos de contraste para abordar el problema dado en (7). El primero compara un estimador no paramétrico con un ajuste paramétrico de la función de regresión:

$$T^{c,1}_{n,p} = \int_{\mathcal{D}} \{1 - \cos[\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p) - m^c_{\hat{\boldsymbol{\beta}}}(\mathbf{x})]\} w(\mathbf{x}) d\mathbf{x}, \quad p = 0, 1,$$

donde w es una función de pesos que ayuda a mitigar el posible efecto frontera. Los estimadores $\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p)$, para $p = 0, 1$, son los estimadores de tipo Nadaraya–Watson o lineal local de la función de regresión circular m^c, dados en (5). Para el estimador paramétrico se consideran procedimientos basados en métodos de mínimos cuadrados y máxima verosimilitud adaptados al contexto de datos circulares (Fisher y Lee, 1992;Lund, 1999;Presnell et al., 1998). El segundo estadístico de contraste que se propone compara un estimador no paramétrico con una versión suavizada del ajuste paramétrico:

$$T^{c,2}_{n,p} = \int_{\mathcal{D}} \{1 - \cos[\hat{m}^c_{\mathbf{H}}(\mathbf{x}; p) - \hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)]\} w(\mathbf{x}) d\mathbf{x}, \quad p = 0, 1, \tag{8}$$

donde $\hat{m}^c_{\mathbf{H},\hat{\boldsymbol{\beta}}}(\mathbf{x}; p)$, $p = 0, 1$, son versiones suavizadas del estimador paramétrico $m^c_{\hat{\boldsymbol{\beta}}}$.

El cálculo de la distribución asintótica de estos estadísticos de contraste no se ha considerado dentro de los objetivos de esta tesis, pero aunque se obtuviera dicha distribución, sería necesario introducir un procedimiento de calibrado en la práctica. La calibración de los valores críticos se realiza empleando métodos bootstrap.

Para datos independientes, se proponen dos métodos bootstrap adaptados al contexto de modelos de regresión con respuesta circular y predictor Euclídeo multidimensional: un bootstrap residual paramétrico circular (PCB) y un bootstrap residual no paramétrico circular (NPCB). En PCB se utilizan los residuos obtenidos del ajuste paramétrico en el algoritmo bootstrap. Si la función de regresión circular pertenece a la familia paramétrica considerada en la hipótesis nula, entonces los residuos tenderán a ser *similares* a los errores teóricos y, por tanto, se espera que el método PCB tenga un buen comportamiento. En NPB, siguiendo las ideas empleadas en González-Manteiga y Cao (1993) para aumentar la potencia del contraste, se usan los residuos obtenidos del ajuste no paramétrico en el procedimiento bootstrap. Estos métodos de remuestreo (PCB y NPCB), diseñados para datos independientes, deben adaptarse adecuadamente en un contexto de dependencia espacial. En este capítulo también se introducen dos procedimientos específicos para calibrar el contraste teniendo en cuenta la estructura de dependencia: un bootstrap residual circular espacial paramétrico (PSCB) y un bootstrap residual circular espacial no paramétrico (NPSCB). Análogamente a PCB, pero en este caso para datos espacialmente correlados, en PSCB se consideran los residuos obtenidos a partir del

ajuste paramétrico bajo la hipótesis nula. La principal diferencia entre PCB y PSCB es que en PSCB, para imitar la estructura de dependencia de los errores, se ajusta un proceso circular espacial a los residuos. En el algoritmo bootstrap se emplean muestras generadas del proceso ajustado. Los pasos seguidos en el procedimiento NPSCB son similares a los empleados en PSCB, pero en el caso de NPSCB, los residuos se obtienen del estimador de regresión no paramétrico.

Se analiza el funcionamiento de los métodos bootstrap propuestos (tanto para datos independientes como para datos espacialmente correlados) a partir de un exhaustivo estudio de simulación. En este estudio se muestra que para datos independientes, en la mayoría de los escenarios considerados, los resultados obtenidos con NPCB son ligeramente mejores a los proporcionados por PCB, especialmente para las hipótesis alternativas consideradas. Además, se observa un comportamiento mejor cuando se emplea $T_{n,1}^{c,2}$, dado en (8). Para datos espacialmente correlados, se obtiene que ambos estadísticos de contraste no funcionan correctamente bajo la hipótesis nula, cuando se utilizan PCB y NPCB, diseñados para datos independientes. Con respecto a PSCB y NPSCB, el uso de residuos no paramétricos en el procedimiento bootstrap proporciona mejores resultados. La potencia del contraste es mayor cuando la estructura de dependencia espacial es más débil. Además, se obtienen resultados más satisfactorios cuando se usa $T_{n,1}^{c,2}$. En ambos contextos (datos independientes y datos espacialmente correlados), las proporciones de rechazos de la hipótesis nula dependen de la matriz ventana $\mathbf{H}$ considerada. El procedimiento de contraste propuesto también se ilustra con dos conjuntos de datos reales: direcciones de bígaros azules (Fisher y Lee, 1992) y direcciones de escape de dos especies de pulgas (Scapini et al., 2002).

Las contribuciones de este capítulo se pueden encontrar en Meilán-Vila et al. (2020c).

Capítulo 6: Conclusiones e investigación futura.

En este capítulo se presentan las conclusiones finales de la tesis y se plantean posibles trabajos futuros.

En primer lugar, en el contexto de modelos de regresión con errores espacialmente correlados, se contrastó si la función de regresión pertenece a una familia paramétrica.

En esta misma línea, se podría calcular la distribución asintótica del estadístico de contraste bajo la suposición de heterocedasticidad de los errores. Para ello, se podrían emplear las ideas consideradas por Fernández-Casal et al. (2017). Además, se podría analizar la consistencia de los algoritmos bootstrap propuestos, por ejemplo, mediante un procedimiento de *imitación* (Shao y Tu, 2012).

Para modelos de regresión con respuesta circular y predictor Euclídeo multidimensional, se propuso un estimador no paramétrico de la función de regresión. Este estimador se estudió también para datos espacialmente correlados. Como línea de trabajo futura, se podrían diseñar nuevos métodos de selección de la matriz ventana en el procedimiento de estimación, por ejemplo, empleando métodos tipo bootstrap. En el contexto de datos espacialmente correlados, sería interesante proponer un procedimiento que permita una elección automática del radio l del círculo en $N_i = \{\mathbf{X}_j : \|\mathbf{X}_j - \mathbf{X}_i\| \leq l\}$, es decir, el número de observaciones que deben eliminarse en la validación cruzada modificada. Además, el estimador propuesto podría estudiarse incluyendo otro tipo de covariables, como categóricas o direccionales, entre otras. Para ello, se podrían emplear las ideas consideradas por García-Portugués et al. (2013), Racine y Li (2004) o Li y Racine (2004).

Por último, se propuso y se analizó empíricamente un contraste de bondad de ajuste para determinar si la función de regresión pertenece a una familia paramétrica en un contexto de modelos con respuesta circular y predictor muldimensional Euclídeo, tanto para datos independientes como para datos espacialmente correlados. En esta parte, se podría calcular la distribución asintótica de los estadísticos de contraste propuesto. Además, para datos espacialmente correlados, en el procedimiento bootstrap, resultaría interesante proponer un modelo que permita ajustar correctamente los residuos circulares con dependencia espacial, y así imitar de forma satisfactoria la estractura de dependencia de los errores.

Apéndice A: Resultados auxiliares.

En este apéndice se incluyen resultados auxiliares que son necesarios para desarrollar las demostraciones de los Capítulos 2, 3 y 4.

Bibliography

Adler, R. J. (2010). *The Geometry of Random Fields*. Wiley, New York.

Alcalá, J., Cristóbal, J., and González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statistics & Probability Letters*, 42:39–46.

Alegría, A., Porcu, E., Furrer, R., and Mateu, J. (2019). Covariance functions for multivariate Gaussian fields evolving temporally over planet earth. *Stochastic Environmental Research and Risk Assessment*, 33(8-9):1593–1608.

Alonso-Pena, M., Ameijeiras-Alonso, J., and Crujeiras, R. M. (2020). Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation,* https://doi.org/10.1080/00949655.2020.1818243.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, 85(411):749–759.

Ameijeiras-Alonso, J., Benali, A., Crujeiras, R. M., Rodríguez-Casal, A., Pereira, J. M., et al. (2019). Fire seasonality identification with multimodality tests. *The Annals of Applied Statistics*, 13(4):2120–2139.

Anastassiou, G. A. (1997). Multivariate ostrowski type inequalities. *Acta Mathematica Hungarica*, 76(4):267–278.

Azzalini, A., Bowman, A. W., and Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika*, 76:1–11.

Batschelet, E. (1981). *Circular Statistics in Biology*. Mathematics in Biology. Academic Press, New York.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika*, 74(3):457–468.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83(403):687–697.

Best, D. J. and Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von mises-fisher concentration parameters: the bias of the maximum likelihood estimators. *Communications in Statistics-Simulation and Computation*, 10(5):493–502.

Biedermann, S. and Dette, H. (2000). Testing linearity of regression models with dependent errors by kernel based methods. *TEST*, 9:417–438.

Bingham, M. and Mardia, K. (1975). Maximum likelihood characterization of the von mises distribution. In *A Modern Course on Statistical Distributions in Scientific Work*, pages 387–398. Springer.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*, volume 18. Oxford Unversity Press, Oxford.

Bowman, A. W. and Crujeiras, R. M. (2013). Inference for variograms. *Computational Statistics & Data Analysis*, 66:19–31.

Carnicero, J. A., Ausín, M. C., and Wiper, M. P. (2013). Non-parametric copulas for circular–linear and circular–circular data: an application to wind directions. *Stochastic Environmental Research and Risk Assessment*, 27(8):1991–2002.

Casson, E. and Coles, S. (1998). Extreme hurricane wind speeds: estimation, extrapolation and spatial smoothing. *Journal of Wind Engineering and Industrial Aerodynamics*, 74:131–140.

Castillo-Páez, S., Fernández-Casal, R., and García-Soidán, P. (2019). A nonparametric bootstrap method for spatial data. *Computational Statistics & Data Analysis*, 137:1–15.

Chiles, J.-P. and Delfiner, P. (2009). *Geostatistics: Modeling Spatial Uncertainty*, volume 497. John Wiley & Sons, New Jersey.

Chiu, S.-T. (1989). Bandwidth selection for kernel estimate with correlated noise. *Statistics & Probability Letters*, 8(4):347–354.

Chu, C.-K., Marron, J. S., et al. (1991). Choosing a kernel regression estimator. *Statistical Science*, 6(4):404–419.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74(368):829–836.

Cox, D. and Lewis, P. (1996). *The Statistical Analysis of Series of Events*. Chapman and Hall, London.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 31:377–403.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology*, 17(5):563–586.

Cressie, N. (1993). *Statistics for Spatial Data*. Wiley, New York.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*, volume 1. Cambridge University Press, Cambridge.

Deschepper, E., Thas, O., and Ottoy, J. P. (2008). Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, 64(3):912–920.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2012). Non-parametric smoothing and prediction for nonlinear circular time series. *Journal of Time Series Analysis*, 33(4):620–630.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40(2):238–255.

Di Marzio, M., Panzera, A., and Taylor, C. C. (2014). Nonparametric regression for spherical data. *Journal of the American Statistical Association*, 109(506):748–763.

Diblasi, A. and Bowman, A. (2001). On the use of the variogram in checking for independence in spatial data. *Biometrics*, 57:211–218.

Diggle, P. and Ribeiro, P. J. (2007). *Model-Based Geostatistics*. Springer, New York.

Diggle, P., Menezes, R., and Su, T. L. (2010). Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59:191–232.

Duong, T. (2015). Spherically symmetric multivariate beta family kernels. *Statistics & Probability Letters*, 104:141–145.

Emery, X. and Porcu, E. (2019). Simulating isotropic vector-valued Gaussian random fields on the sphere through finite harmonics approximations. *Stochastic Environmental Research and Risk Assessment*, 33(8-9):1659–1667.

Eubank, R. L., Li, C. S., and Wang, S. (2005). Testing lack-of-fit of parametric regression models using nonparametric regression techniques. *Statistica Sinica*, 15:135–152.

Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Association*, 85:387–392.

Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association*, 87(420):998–1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, 21(1):196–216.

Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engel, J. (1997). Local polynomial regression: optimal kernels and asymptotic minimax efficiency. *Annals of the Institute of Statistical Mathematics*, 49(1):79–99.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20(4):2008–2036.

Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 57(2):371–394.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman and Hall, London.

Fan, J., Gijbels, I., Hu, T.-C., and Huang, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica*, 6:113–127.

Fan, J., Zhang, C., and Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics*, 29:153–193.

Fernández-Casal, R. (2019). *npsp: Nonparametric spatial (geo)statistics*. R package version 0.7-5.

Fernández-Casal, R., Castillo-Páez, S., and García-Soidán, P. (2017). Nonparametric estimation of the small-scale variability of heteroscedastic spatial processes. *Spatial Statistics*, 22:358–370.

Fernández-Casal, R. and Francisco-Fernández, M. (2014). Nonparametric bias-corrected variogram estimation under non-constant trend. *Stochastic Environmental Research and Risk Assessment*, 28(5):1247–1259.

Fernández-Casal, R., González-Manteiga, W., and Febrero-Bande, M. (2003a). Flexible spatio-temporal stationary variogram models. *Statistics and Computing*, 13(2):127–136.

Fernández-Casal, R., González-Manteiga, W., and Febrero-Bande, M. (2003b). Space-time dependency modeling using general classes of flexible stationary var-

iogram models. *Journal of Geophysical Research: Atmospheres*, 108(D24):1771–1796.

Fisher, N. (1995). *Statistical Analysis of Circular Data*. Cambridge University Press, Cambridge.

Fisher, N. and Lee, A. (1992). Regression models for an angular response. *Biometrics*, 48(3):665–677.

Fisher, N. and Lee, A. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 56(2):327–339.

Francisco-Fernández, M. and Opsomer, J. D. (2005). Smoothing parameter selection methods for nonparametric regression with spatially correlated errors. *Canadian Journal of Statistics*, 33:279–295.

Francisco-Fernández, M., Quintela-del Río, A., and Fernández-Casal, R. (2012). Nonparametric methods for spatial regression. An application to seismic events. *Environmetrics*, 23(1):85–93.

Francisco-Fernández, M. and Vilar-Fernández, J. M. (2001). Local polynomial regression estimation with correlated errors. *Communications in Statistics-Theory and Methods*, 30(7):1271–1293.

García-Portugués, E., Barros, A. M., Crujeiras, R. M., González-Manteiga, W., and Pereira, J. (2014). A test for directional-linear independence, with applications to wildfire orientation and size. *Stochastic Environmental Research and Risk Assessment*, 28(5):1261–1275.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Kernel density estimation for directional–linear data. *Journal of Multivariate Analysis*, 121:152–175.

García-Portugués, E., Van Keilegom, I., Crujeiras and, R. M., and González-Manteiga, W. (2016). Testing parametric models in linear-directional regression. *Scandinavian Journal of Statistics*, 43(4):1178–1191.

Gasser, T. and Müller, H. G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation*, volume 757, pages 23–68. Springer.

González-Manteiga, W. and Cao, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *TEST*, 2(1-2):161–188.

González-Manteiga, W. and Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *TEST*, 22(3):361–411.

González-Manteiga, W., Martinez-Miranda, M., and Pérez-González, A. (2004). The choice of smoothing parameter in nonparametric regression through wild bootstrap. *Computational Statistics & Data Analysis*, 47(3):487–515.

González Manteiga, W. and Vilar Fernández, J. M. (1995). Testing linear regression models using non-parametric regression estimators when errors are non-independent. *Computational Statistics & Data Analysis*, 20(5):521–541.

Gu, J., Li, Q., and Yang, J. (2015). Multivariate local polynomial kernel estimators: Leading bias and asymptotic distribution. *Econometric Reviews*, 34(6-10):979–1010.

Hall, P. (1986). On the bootstrap and confidence intervals. *The Annals of Statistics*, 14:1431–1452.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *The Annals of Statistics*, 20(2):675–694.

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, 99(3):399–424.

Hall, P. and Robinson, A. P. (2009). Reducing variability of crossvalidation for smoothing parameter choice. *Biometrika*, 96:175–186.

Hallin, M., Lu, Z., and Tran, L. T. (2004). Local linear spatial regression. *The Annals of Statistics*, 32:2469–2500.

Hamelryck, T., Mardia, K., and Ferkinghoff-Borg, J. (2012). *Bayesian Methods in Structural Bioinformatics*. Springer, New York.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21:1926–1947.

Härdle, W. and Müller, M. (2012). Multivariate and semiparametric kernel regression. In *Smoothing and Regression: Approaches, Computation, and Application*. John Wiley & Sons, New Jersey.

Härdle, W. and Vieu, P. (1992). Kernel regression smoothing of time series. *Journal of Time Series Analysis*, 13(3):209–232.

Harper, W. V. and Furr, J. M. (1986). Geostatistical analysis of potentiometric data in Wolfcamp aquifer of the Palo Duro Basin, Texas. Technical report, Battelle Memorial Inst.

Hart, J. D. (2013). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York.

Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 53(1):173–187.

Hart, J. D. (1994). Automated kernel smoothing of dependent data by using time series cross-validation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 56(3):529–542.

Hart, J. D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *The Annals of Statistics*, 18(2):873–890.

Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science*, 8(2):120–129.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109.

Jammalamadaka, S. R. and Sarma, Y. R. (1988). A correlation coefficient for angular variables. In Matusita, K., editor, *Statistical Theory and Data Analysis II*, pages 349–364. Amsterdam North-Holland.

Jammalamadaka, S. R. and Sarma, Y. R. (1993). Circular regression. In Matusita, K., editor, *Statistical Science and Data Analysis*, pages 109–128, Utrecht. VSP.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*, volume 5. World Scientific, Singapore.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73(363):602–606.

Jona-Lasinio, G., Gelfand, A., and Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics*, 6(4):1478–1498.

Jona-Lasinio, G., Mastrantonio, G., and Santoro, M. (2019). *CircSpaceTime: Spatial and Spatio-Temporal Bayesian Model for Circular Data*. R package version 0.9.0.

Kim, S. and SenGupta, A. (2017). Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation*, 87(7):1277–1291.

Kim, T. Y., Ha, J., Hwang, S. Y., Park, C., and Luo, Z.-M. (2013). Central limit theorems for reduced U-statistics under dependence and their usefulness. *Australian & New Zealand Journal of Statistics*, 55:387–399.

Köhler, M., Schindler, A., and Sperlich, S. (2014). A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, 82(2):243–274.

Kozek, A. S. (1991). A nonparametric test of fit of a parametric model. *Journal of Multivariate Analysis*, 37(1):66–75.

Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147.

Lagona, F., Picone, M., and Maruotti, A. (2015). A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, 26(8):534–544.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer, New York.

Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, 14(4):457–471.

Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. Chapman and Hall/CRC, Boca Ratón.

Li, C. S. (2005). Using local linear kernel smoothers to test the lack of fit of nonlinear regression models. *Statistical Methodology*, 2(4):267–284.

Li, Q. and Racine, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 14(2):485–512.

Liu, X. H. (2001). *Kernel smoothing for spatially correlated data*. PhD thesis, Department of Statistics, Iowa State University.

Lund, U. (1999). Least circular distance regression for directional data. *Journal of Applied Statistics*, 26(6):723–733.

Lund, U., Agostinelli, C., Arai, H., Gagliardi, A., Garcia-Portugués, E., Giunchi, D., Irisson, J. O., Pocernich, M., and Rotolo, F. (2020). *circular: Circular Statistics*. R package version 0.4-93.

Maglione, D. and Diblasi, A. (2004). Exploring a valid model for the variogram of an isotropic spatial process. *Stochastic Environmental Research and Risk Assessment*, 18:366–376.

Marchetti, G. M. and Scapini, F. (2003). Use of multiple regression models in the study of sandhopper orientation under natural conditions. *Estuarine, Coastal and Shelf Science*, 58:207–215.

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. Wiley, Chichester.

Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, 24(2):165–179.

Mastrantonio, G., Gelfand, A. E., and Lasinio, G. J. (2016). The wrapped skew Gaussian process for analyzing spatio-temporal data. *Stochastic Environmental Research and Risk Assessment*, 30(8):2231–2242.

Mastrantonio, G., Pollice, A., and Fedele, F. (2018). Distributions-oriented wind forecast verification by a hidden Markov model for multivariate circular–linear data. *Stochastic Environmental Research and Risk Assessment*, 32(1):169–181.

Matheron, G. (1962). Traité de géostatistique appliquée, tome i: Mémoires du bureau de recherches géologiques et minières. *Editions Technip, Paris*, 14.

Meilán-Vila, A., Crujeiras, R. M., and Francisco-Fernández, M. (2020a). Nonparametric estimation of circular trend surfaces with application to wave directions. *Stochastic Environmental Research and Risk Assessment,* https://doi.org/10.1007/s00477-020-01919-5.

Meilán-Vila, A., Fernández-Casal, R., Crujeiras, R. M., and Francisco-Fernández, M. (2020b). A computational validation for nonparametric assessment of spatial trends. *arXiv*, 2002.05489. Submitted.

Meilán-Vila, A., Francisco-Fernández, M., and Crujeiras, R. M. (2020c). Goodness-of-fit tests for parametric regression models with circular response. *arXiv*, 2008.13473. Submitted.

Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R. M., and Panzera, A. (2020d). Nonparametric multiple regression estimation for circular response. *TEST,* https://doi.org/10.1007/s11749-020-00736-w.

Meilán-Vila, A., Opsomer, J. D., Francisco-Fernández, M., and Crujeiras, R. M. (2020e). A goodness-of-fit test for regression models with spatially correlated errors. *TEST*, 29:728–749.

Modlin, D., Fuentes, M., and Reich, B. (2012). Circular conditional autoregressive modeling of vector fields. *Environmetrics*, 23(1):46–53.

Morphet, W. J. (2009). *Simulation, kriging, and visualization of circular-spatial data*. PhD thesis, Utah State University.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, 9:141–142.

Neuman, S. P. and Jacobson, E. A. (1984). Analysis of nonintrinsic spatial variability by residual kriging with application to regional groundwater levels. *Journal of the International Association for Mathematical Geology*, 16(5):499–521.

Olea, R. A. and Pardo-Iguzquiza, E. (2011). Generalized bootstrap method for assessment of uncertainty in semivariogram inference. *Mathematical Geosciences*, 43(2):203–228.

Opsomer, J. and Francisco-Fernández, M. (2010). Finding local departures from a parametric model using nonparametric regression. *Statistical Papers*, 51(1):69–84.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16:134–153.

Papadakis, M., Tsagris, M., Dimitriadis, M., Fafalios, S., Tsamardinos, I., Fasiolo, M., Borboudakis, G., Burkardt, J., Zou, C., Lakiotaki, K., and Chatzipantsiou, C. (2020). *Rfast: A Collection of Efficient and Extremely Fast R Functions*. R package version 2.0.1.

Park, C., Kim, T. Y., Ha, J., Luo, Z.-M., and Hwang, S. Y. (2015). Using a bimodal kernel for a nonparametric regression specification test. *Statistica Sinica*, 25:1145–1161.

Porcu, E., Bevilacqua, M., and Genton, M. G. (2016). Spatio-temporal covariance and cross-covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association*, 111(514):888–898.

Presnell, B., Morrison, S. P., and Littell, R. C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, 93(443):1068–1077.

Priestley, M. B. and Chao, M. (1972). Non-parametric function fitting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 34(3):385–392.

R Development Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Racine, J. and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, 119(1):99–130.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. Wiley, New York.

Ribeiro, P. J. and Diggle, P. J. (2020). *geoR: Analysis of geostatistical data*. R package version 1.7-5.2.2.

Rivest, L.-P., Duchesne, T., Nicosia, A., and Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65(3):445–463.

Rozanov, Y. A. (1967). *Stationary Random Processes*. Holden-Day, Oakland.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, 22:1346–1370.

Scapini, F., Aloia, A., Bouslama, M. F., Chelazzi, L., Colombini, I., ElGtari, M., Fallaci, M., and Marchetti, G. M. (2002). Multiple regression analysis of the sources of variation in orientation of two sympatric sandhoppers, *Talitrus saltator* and *Talorchestia brito*, from an exposed mediterranean beach. *Behavioral Ecology and Sociobiology*, 51(5):403–414.

Schmidt-Koenig, K. (1963). On the role of the loft, the distance and site of release in pigeon homing (the "cross-loft experiment"). *The Biological Bulletin*, 125(1):154–164.

SenGupta, A. and Ugwuowo, F. I. (2006). Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environmental and Ecological Statistics*, 13(3):299–309.

Shao, J. and Tu, D. (2012). *The Jackknife and Bootstrap*. Springer, New York.

Shapiro, A. and Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, 11(1):87–96.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*, volume 26. Chapman and Hall, London.

Solow, A. R. (1985). Bootstrapping correlated data. *Mathematical Geology*, 17(7):769–775.

Sperlich, S. (2013). Comments on: An updated review of goodness-of-fit tests for regression models. *TEST*, 22(3):419–427.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 5(4):595–620.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, 25:613–641.

Székely, G. J., Rizzo, M. L., Bakirov, N. K., et al. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6):2769–2794.

Vilar-Fernández, J. M. and Francisco-Fernández, M. (2002). Local polynomial regression smoothers with AR-error structure. *TEST*, 11(2):439–464.

Vilar-Fernández, J. M. and González-Manteiga, W. (1996). Bootstrap test of goodness of fit to a linear model when errors are correlated. *Communications in Statistics Theory and Methods*, 25:2925–2953.

Wang, F. and Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. *Statistical Methodology*, 10(1):113–127.

Wang, F. and Gelfand, A. E. (2014). Modeling space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association*, 109(508):1565–1580.

Wang, F., Gelfand, A. E., and Jona-Lasinio, G. (2015). Joint spatio-temporal analysis of a linear and a directional variable: space-time modeling of wave heights and wave directions in the adriatic sea. *Statistica Sinica*, 25:25–39.

Watson, G. S. (1964). Smooth regression analysis. *Sankhya*, 26:359–372.

Weihrather, G. (1993). Testing a linear regression model against nonparametric alternatives. *Metrika*, 40:367–379.

Yang, L. and Tschernig, R. (1999). Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(4):793–815.